- **阅读网络文献，了解什么是 Gibbs Sampling，并知道如何实现。**

为什么要用吉布斯采样

什么是 sampling?

sampling 就是以一定的概率分布，看发生什么事件。举一个例子。甲只能 E：吃饭、学习、打球，时间 T：上午、下午、晚上，天气 W：晴朗、刮风、下雨。现在要一个 sample，这个 sample 可以是：打球+下午+晴朗。

吉布斯采样的通俗解释？

问题是我们不知道 p(E,T,W)，或者说，不知道三件事的联合分布 joint distribution。当然，如果知道的话，就没有必要用 gibbs sampling 了。但是，我们知道三件事的 conditional distribution。也就是说，p(E|T,W),p(T|E,W),p(W|E,T)。现在要做的就是通过这三个已知的条件分布，再用 gibbs sampling 的方法，得到联合分布。

具体方法。首先随便初始化一个组合,i.e. 学习+晚上+刮风，然后依条件概率改变其中的一个变量。具体说，假设我们知道晚上+刮风，我们给 E 生成一个变量，比如，学习-》吃饭。我们再依条件概率改下一个变量，根据学习+刮风，把晚上变成上午。类似地，把刮风变成刮风（当然可以变成相同的变量）。这样学习+晚上+刮风-》吃饭+上午+刮风。同样的方法，得到一个序列，每个单元包含三个变量，也就是一个马尔可夫链。然后跳过初始的一定数量的单元（比如 100 个），然后隔一定的数量取一个单元（比如隔 20 个取 1 个）。这样 sample 到的单元，是逼近联合分布的。

14.11 Consider the query $\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$ in Figure 14.11(a) and how MCMC can answer it.

a. How many states does the Markov chain have?

b. Calculate the **transition matrix** Q containing $q(y \rightarrow y')$ for all y, y'.

c. What does $\mathbf{Q}^2$, the square of the transition matrix, represent?

d. What about Q^n as $n \rightarrow \infty$?

e. Explain how to do probabilistic inference in Bayesian networks, assuming that Q^n is available. Is this a practical way to do inference?

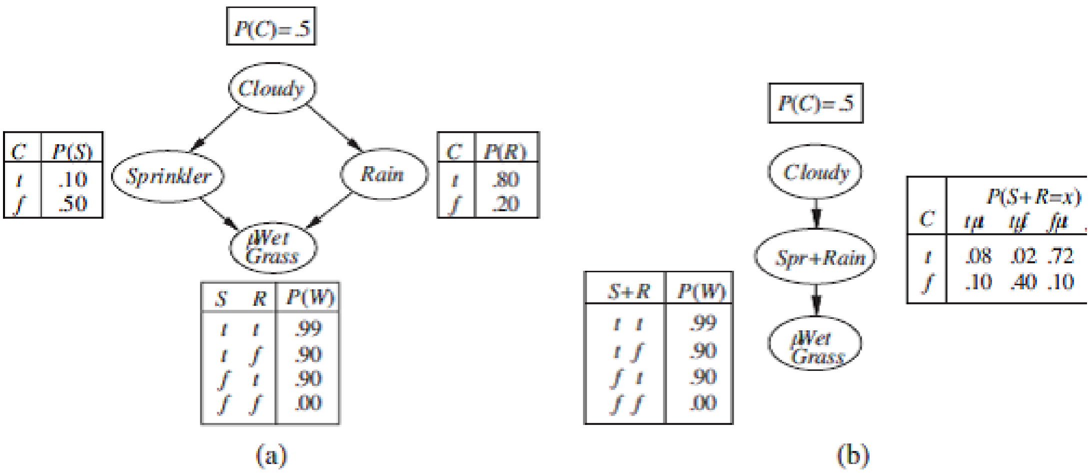

Figure 14.12 (a) A multiply connected network with conditional probability tables. (b) A clustered equivalent of the multiply connected network.

a．The Markov Chain have 4 states. There are two non-evidence variables Cloudy and Rain. Since each variable can have two states true and flase, there will be four states for the two variables.

a．马尔可夫链有 4 种状态。有两个隐变量阴天和下雨。因为每个变量可以有两个状态：真和 FLASH，这两个变量将有四个状态。

b.

By computing the sampling distributions for each variable, the transition matrix can be obtained. Sampling distributions for each variable are:

$$P(C \mid r,s) = \alpha P(C)P(s \mid C)P(r \mid C)$$
$$= \alpha \langle 0.5,0.5\rangle \langle 0.1,0.5\rangle \langle 0.8,0.2\rangle$$
$$= \alpha \langle 0.04,0.05\rangle$$
$$= \langle 4/9,5/9\rangle$$

$$P(C \mid \neg r,s) = \alpha P(C)P(s \mid C)P(\neg r \mid C)$$
$$= \alpha \langle 0.5,0.5\rangle \langle 0.1,0.5\rangle \langle 0.2,0.8\rangle$$
$$= \alpha \langle 0.01,0.20\rangle$$
$$= \langle 1/21,20/21\rangle$$

$$P(R \mid c,s,w) = \alpha P(R \mid c)P(w \mid s,R)$$
$$= \alpha \langle 0.8,0.2\rangle \langle 0.99,0.90\rangle$$
$$= \alpha \langle 0.792,0.180\rangle$$
$$= \langle 22/27,5/27\rangle$$

c.

It represent the probability of reaching a state and the superscript 2 repersents that two steps are taken to move from one state to anther state.

d.

When n→∞， the probability of being n each state reaches infinity. That means being in each state will have longer-probability.

e

The process of computing the posterior distribution of variables given evidence is called probabilistic inference . In Bayesian networks, when Q^n is available the probabilistic inference can be obtained by simpler multiplications.

For example if Q^I is available , Q^2 can be obtained with one multiplication. But this is not a practical way to do inference because in a Bayesian network with n Boolean variables, the matrix will be of size $2^n * 2^n$ and each multiplication operation takes O(2^{3n}) operations.