Project nr 2

Gibbs sampling

Mateusz Kapusta

January 15, 2023

1 Rain BBN

P(C)
0.5

C	P(S|C)
T	0.1
F	0.5

C	P(R|C)
T	0.8
F	0.2

C	R	P(W|S,R)
T	**T**	0.99
T	**F**	0.9
F	**T**	0.9
F	**F**	0.01

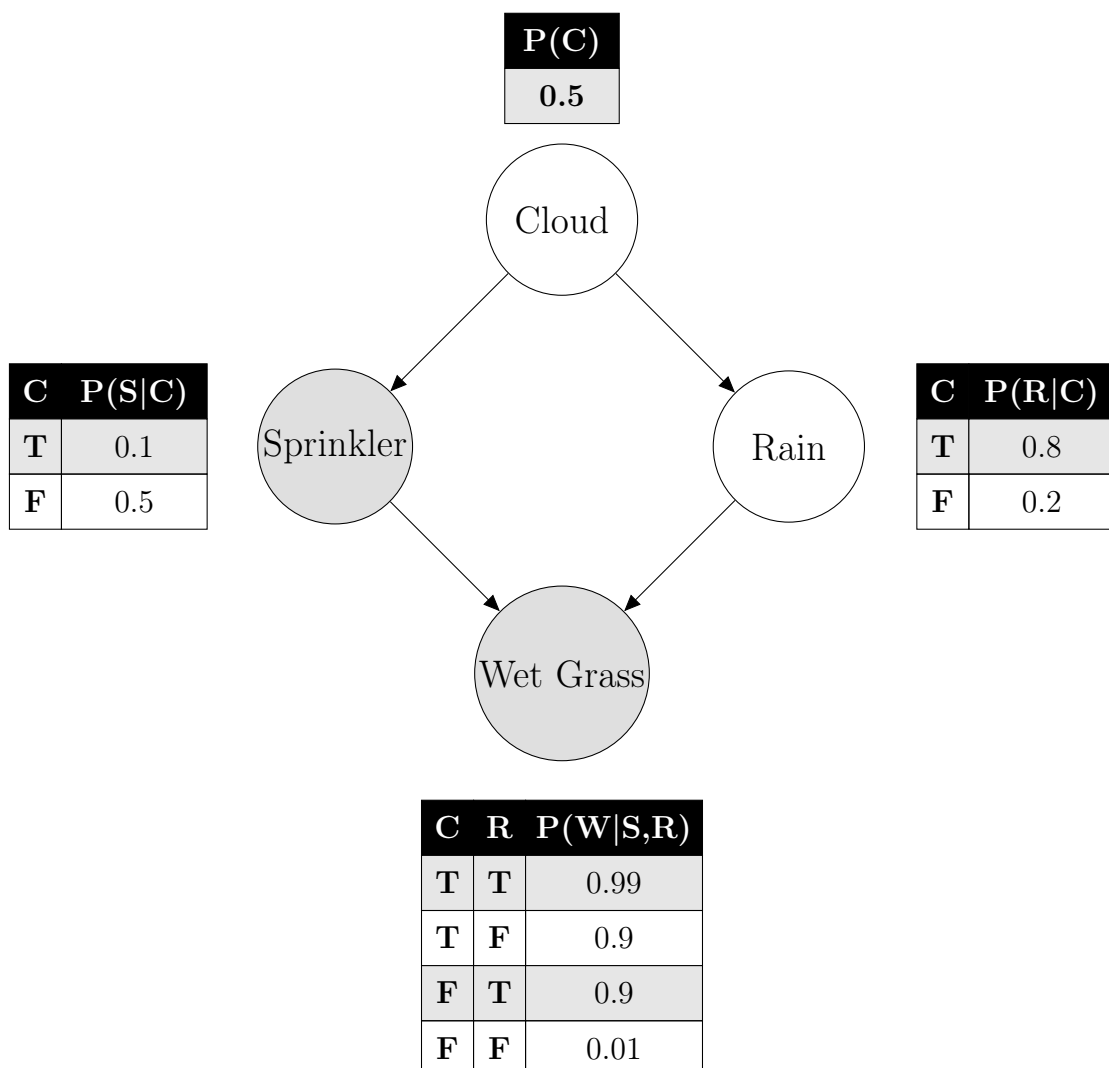Figure 1: Bayesian diagram for rain network

Let's conisder Bayesian Belif network depicted at 1. Each of the varibles can be either True or False. Conditional dependence between them in encoded in the matrix next to each of the nodes. Right now let's do not think about which varibles are observed and try to dervie formula for propability of one node with respect to other varibles which remain constant

1. $P(C = T | R = T, S = T, W = T)$

   We can decompose local propability distribution to

   $$P(C = T | R = T, S = T, W = T) \propto P(S = T | C = T)P(R = T | C = T) = 0.1 \cdot 0.8 = 0.08$$

   In order to find true propability we need to compute renormalazing constant

   $$P(C = F | R = T, S = T, W = T) \propto P(S = T | C = F)P(R = T | C = F) = 0.5 \cdot 0.2 = 0.10$$

   . After renomalization

   $$P(C = T | R = T, S = T, W = T) = \frac{0.08}{0.08 + 0.1} \approx 0.44$$

2. $P(C = T | R = F, S = T, W = T)$

   As before

   $$P(C = T | R = F, S = T, W = T) \propto P(S = T | C = T)P(R = F | C = T) = 0.1 \cdot 0.2 = 0.02$$

   In order to find true propability we need to compute renormalazing constant

   $$P(C = F | R = F, S = T, W = T) \propto P(S = T | C = F)P(R = F | C = F) = 0.5 \cdot 0.8 = 0.4$$

   . After renomalization

   $$P(C = T | R = F, S = T, W = T) = \frac{0.02}{0.02 + 0.4} \approx 0.047$$

3. $P(R = T | C = T, S = T, W = T)$

As before

$$P(R = T|C = T, S = T, W = T) \propto P(W = T|R = T, S = T)P(R = T|C = T) = 0.99 \cdot 0.8 = 0.792$$

In order to find true propability we need to compute renormalazing constant

$$P(R = F|C = T, S = T, W = T) \propto P(W = T|R = F, S = T)P(R = F|C = T) = 0.9 \cdot 0.2 = 0.18$$

. After renomalization

$$P(R = T|C = T, S = T, W = T) = \frac{0.792}{0.18 + 0.792} \approx 0.81$$

4. $P(R = T|C = F, S = T, W = T)$

   As before

$$P(R = T|C = F, S = T, W = T) \propto P(W = T|R = T, S = T)P(R = T|C = F) = 0.99 \cdot 0.2 = 0.198$$

In order to find true propability we need to compute renormalazing constant

$$P(R = F|C = F, S = T, W = T) \propto P(W = T|R = F, S = T)P(R = F|C = F) = 0.9 \cdot 0.8 = 0.72$$

. After renomalization

$$P(R = T|C = F, S = T, W = T) = \frac{0.198}{0.72 + 0.198} \approx 0.22$$

Our procedure is very simple and allows us to determine marbinal propability of one varible with respect to others. This simple procedure allows us to construct Gibbs sampler which uses this simple procedure to sample from our model in general case and will help to estimate what is marginal propability of $P(R = T, S = T, W = T)$.

2   Gibbs sampler

Gibbs sampler was implemented in Python programing language using NumPy. After implemen-

tation 100 samples were drown from model. This small sample allowed to estiamte that $P(R = T, S = T, W = T) \approx 0.18$. This result probably does not represent correct value as there might be a problem with convergence of our chain. Hence we need to fin wheter is it safe to infer from our chain.

3   Convergence Diagnostics

In order to determine wheter our chain in safe to infer from chain two test were performed. Firstly relative frequency of occurence was checked and then autocorrelation test was performed. During the first $N = 50000$ samples were drown from one chain and then realative frequency for each varible denoted as θ was calculated using

$$f(t) = \frac{1}{t} \sum_{0}^{t} \theta_t$$

allowing to present, how our estimates change acros time. Results are presented on figure 2
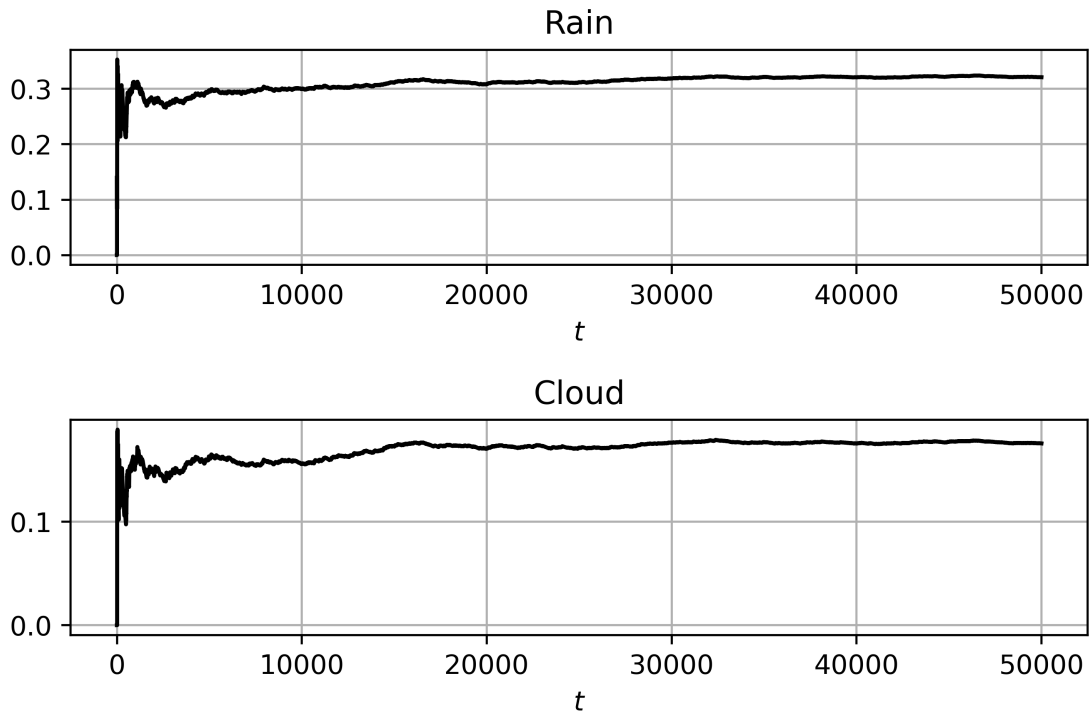

Figure 2: Relative frequency f of parameter against position in chain t for Rain varible and Cloud varible

    We can see, that around $t \approx 10000$ our chain is begining to stabilize and from this point onward our estimates do not change. That need to be emphasized is huge bump at the begining of our sequence. This behaviour suggest our chain is very noisy at the beginning and we should not infer from the

first 100 samples only. Second test was based on the lagged autocorrelation function. Firstly we can introduce lagged sequence $\theta_t^l = \theta_{t+l}$. Due to finite length of our chain last l elements of orignal chain are discarded in this calculation. We can then introduce Pearson corelation coofcient $P(l)$ using equation

$$P(l) = \frac{\sum_{i=0}^{n}(\theta_i - \hat{\theta})(\theta_i^l - \hat{\theta}^l)}{\sqrt{\sum_{i=0}^{n}(\theta_i^2 - (\hat{\theta})^2)}\sqrt{\sum_{i=0}^{n}((\theta_i^l)^2 - (\hat{\theta}^l)^2)}}$$

where $\hat{\theta}$ is mean value of θ while n is effective length of our sample equal to $n = N - l$ (as we discard last l draws). Values of our coofcient were computed for $l \in \langle 0, 100 \rangle$ were computed and ploted on figure 3.
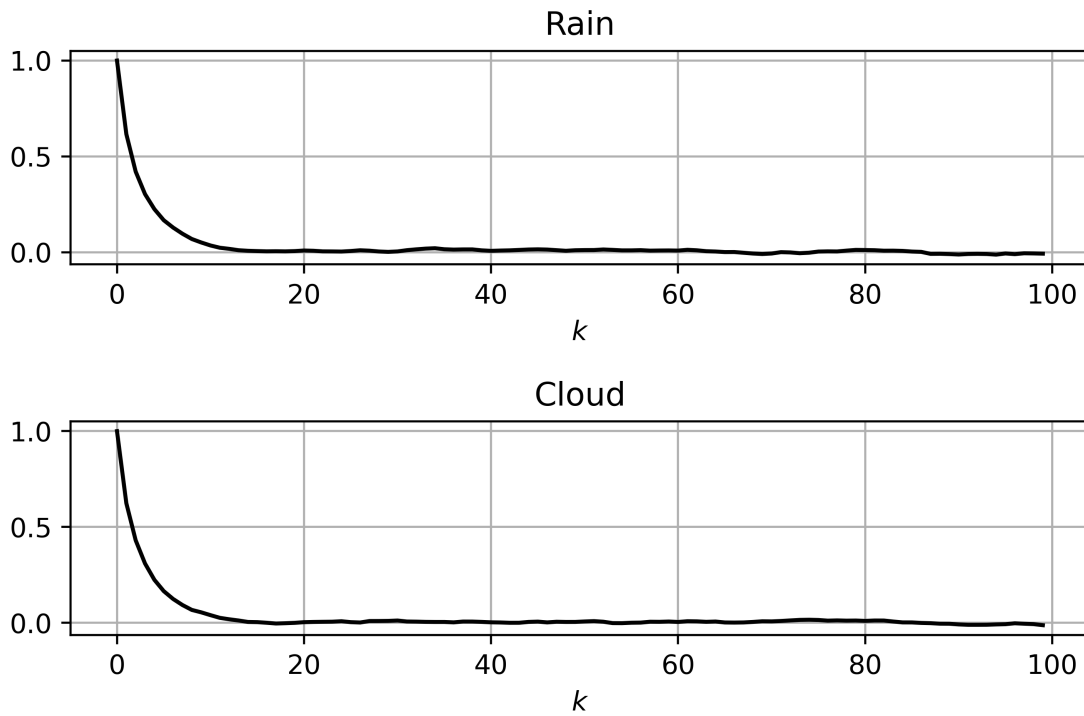

Figure 3: Corelation coofcients $P(l)$ of parameter against lag l for Rain varible and Cloud varible

We can see, that for $l > 0$ autocorrelation is neglectable so suggested thinning out should be around this value.

3.1   Final model

After determining suitable parameters in previous section final sampling took place with burnin $N_b = 10000$ and thining out $l = 50$. Based on MCMC estimation $P(R = T, S = T, W = T) \approx 0.33$ was obtained. Now we can compare this result with exact one. In order to calculate analyticly value of

propability we will follow procedure from first section. We can write

$$P(R = T|S = T, W = T) \propto P(R = T|C = T)P(W = T|R = T, S = T)P(S = T|C = T)+$$

$$+P(R = T|C = F)P(W = T|R = T, S = T)P(S = T|C = F) = 0.8 \cdot 0.99 \cdot 0.1 + 0.2 \cdot 0.99 \cdot 0.5 = 0.1881$$

. On the other hand

$$P(R = F|S = T, W = T) \propto P(R = F|C = T)P(W = T|R = F, S = T)P(S = T|C = T)+$$

$$+P(R = F|C = F)P(W = T|R = F, S = T)P(S = T|C = F) = 0.2 \cdot 0.9 \cdot 0.1 + 0.8 \cdot 0.9 \cdot 0.5 = 0.378$$

. Combining both results we get

$$P(R = T|S = T, W = T) = \frac{0.1881}{0.1881 + 0.378} \approx 0.332$$

. We can hence see that our result is just little of from the exact value. moreover we can see that our estimation is much better then one based on first 100 samples which do not have anything in common with exact result.

3.2   Gelman-Rubin Test

Finaly in order to find wheter chain have reached convergence Gelman-Rubin test was implemented. In order to test for convergence following procedure was followed. M runs of same MCMC model was performed. For each chain mean and variance were calculated

$$\hat{\theta^m} = \frac{1}{M} \sum_{i=1}^{N} \theta_i^m \tag{1}$$

$$\sigma^m = \frac{1}{N-1} \sum_{i=0}^{N} (\theta_i^m - \hat{\theta^m}) \tag{2}$$

. Those parameters present parameters inside each of chains. Then we define mean of all chains

$$\hat{\theta} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta^m}$$

and variance of means across common mean

$$B = \frac{N}{M-1} \sum_{m=1} (\hat{\theta^m} - \hat{\theta})^2$$

. At the end we can introduce avarage variance of all chains

$$W = \frac{1}{M} \sum_{m=0}^{M} \sigma^m$$

together with estimatior

$$\hat{V} = \frac{N-1}{N} W + \frac{M+1}{MN} B$$

. Under hypothesis that we reached convergence $\hat{V}$ is unbiased estimator of true variance. On the other hande in this scenario also W should be unbiased estimator of variance. Hence we can introduce value $R = \sqrt{\frac{\hat{V}}{W}}$ which should be 1 in the case of convergence. Rubin Gelman test was performed using $M = 40$ and for both final sampler parameters and first ones. For 100 samples with no burnout and thinning out statistics for Rain and Cloud node were $R_{rain} \approx 1.117$ and $R_{cloud} \approx 1.116$ respectivly. For final paramters with $N = 100$, burnin $N_b = 10000$ and thining out $l = 50$ statistic had value $R_{rain} = 1.047$ and $R_{cloud} = 1.055$ so we can clearly see we improved in the convergence and and first sample was certainly biased and should not be used to infer parameters.