SDA - Project 2: GIbbs Sampler

Kacper Domżał 386308

1. Derive the formulas for P(C = T | R = T, S = T, W = T), P(C = T | R = F, S = T, W = T), P(R = T | C = T, S = T, W = T) and P(R = T | C = F, S = T, W = T) up to normalization constants in the denominators, and compute their values by renormalizing the two possible values for each conditional probability distribution.

$$P(C=T|R=T,S=T,W=T) = P(C=T|R=T,S=T)$$
$$= \frac{P(C=T)P(R=T|C=T)P(S=T|C=T)}{P(C=T)P(R=T|C=T)P(S=T|C=T)+P(C=F)P(R=T|C=F)P(S=T|C=F)} = \frac{0.5*0.8*0.1}{0.5*0.8*0.1+0.5*0.2*0.5}$$
$$= 0.444$$

$$P(C=F|R=T,S=T,W=T) = 1 - P(C=T|R=T,S=T,W=T) = 0.556$$

$$P(C=T|R=F,S=T,W=T) = P(C=T|R=F,S=T)$$
$$= \frac{P(C=T)P(R=F|C=T)P(S=T|C=T)}{P(C=T)P(R=F|C=T)P(S=T|C=T)+P(C=F)P(R=F|C=F)P(S=T|C=F)} = \frac{0.5*0.2*0.1}{0.5*0.2*0.1+0.5*0.8*0.5}$$
$$= 0.048$$

$$P(C=F|R=F,S=T,W=T) = 1 - P(C=T|R=F,S=T,W=T) = 0.952$$

$$P(R=T|C=T,S=T,W=T) = \frac{P(R=T|C=T,S=T)P(W=T|R=T,C=T,S=T)}{P(W=T|C=T,S=T)}$$

$$= \frac{P(R=T|C=T)P(W=T|R=T,S=T)}{P(R=T|C=T)P(W=T|R=T,S=T)+P(R=F|C=T)P(W=T,R=F,S=T)} = \frac{0.8*0.99}{0.8*0.99+0.2*0.9} = 0.815$$

$$P(R=F|C=T,S=T,W=T) = 1 - P(R=T|C=T,S=T,W=T) = 0.185$$

$$P(R=T|C=F,S=T,W=T) = \frac{P(R=T|C=F,S=T)P(W=T|R=T,C=F,S=T)}{P(W=T|C=F,S=T)}$$

$$= \frac{P(R=T|C=F)P(W=T|R=T,S=T)}{P(R=T|C=F)P(W=T|R=T,S=T)+P(R=F|C=F)P(W=T|R=F,S=T)} = \frac{0.2*0.99}{0.2*0.99+0.8*0.9} = 0.216$$

$$P(R=F|C=F,S=T,W=T) = 1 - P(R=T|C=F,S=T,W=T) = 0.784$$

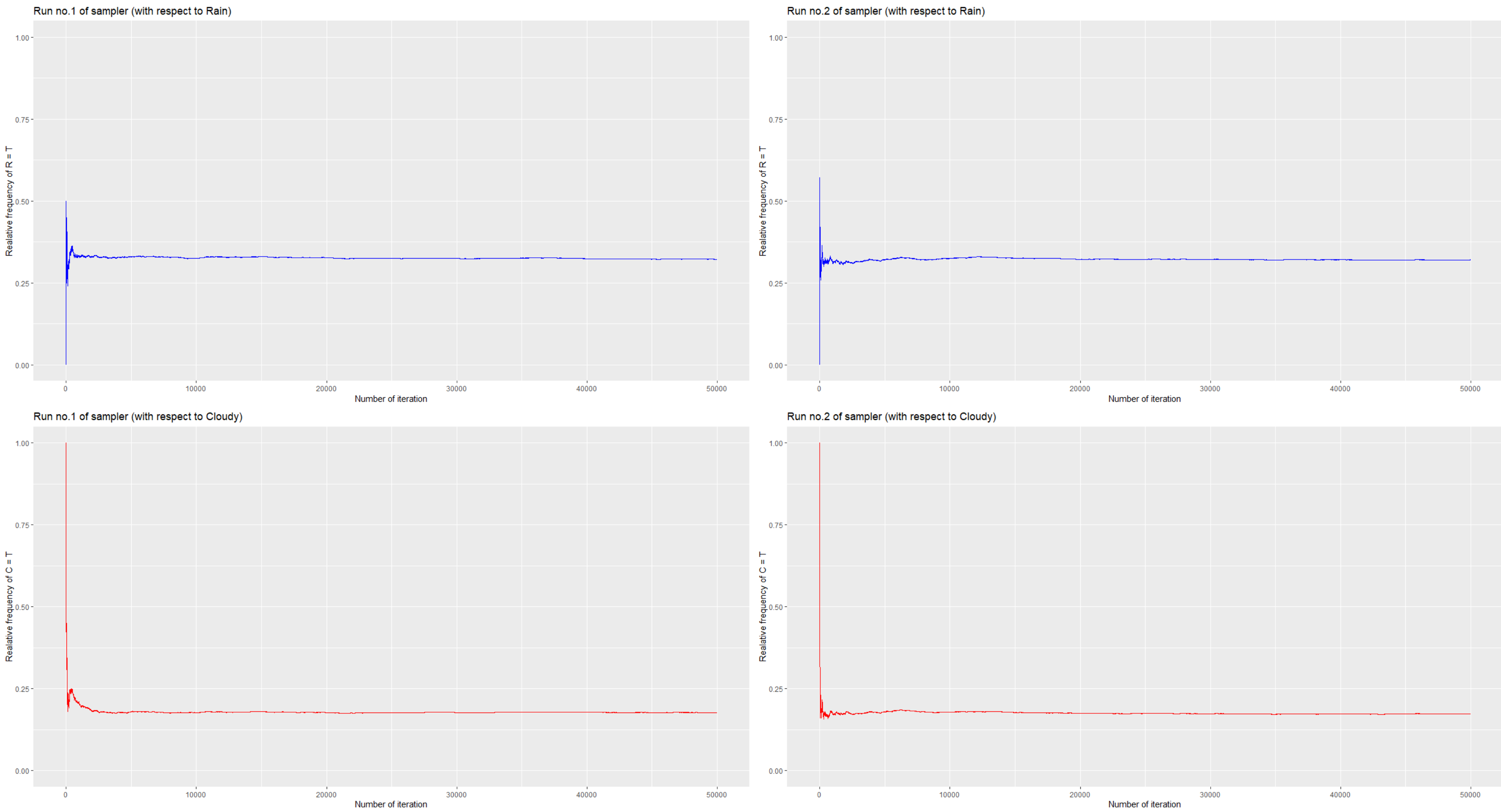
C	R	Probability	R	C	Probability
0	0	0.952	0	0	0.784
0	1	0.556	1	0	0.216
1	0	0.048	0	1	0.185
1	1	0.444	1	1	0.815

2&3. Implement the Gibbs sampler sketched above for the Bayesian network in Figure 1 and draw 100 samples from the joint probability distribution P(R, C | S = T, W = T). Estimate the marginal probability of rain, given that the sprinkler is on and the grass is wet P(R = T | S = T, W = T) from the 100 samples. (1 point).

The estimated marginal probability (from 100 samples) of rain, given that the sprinkler is on and the grass is wet is equal to 0.29

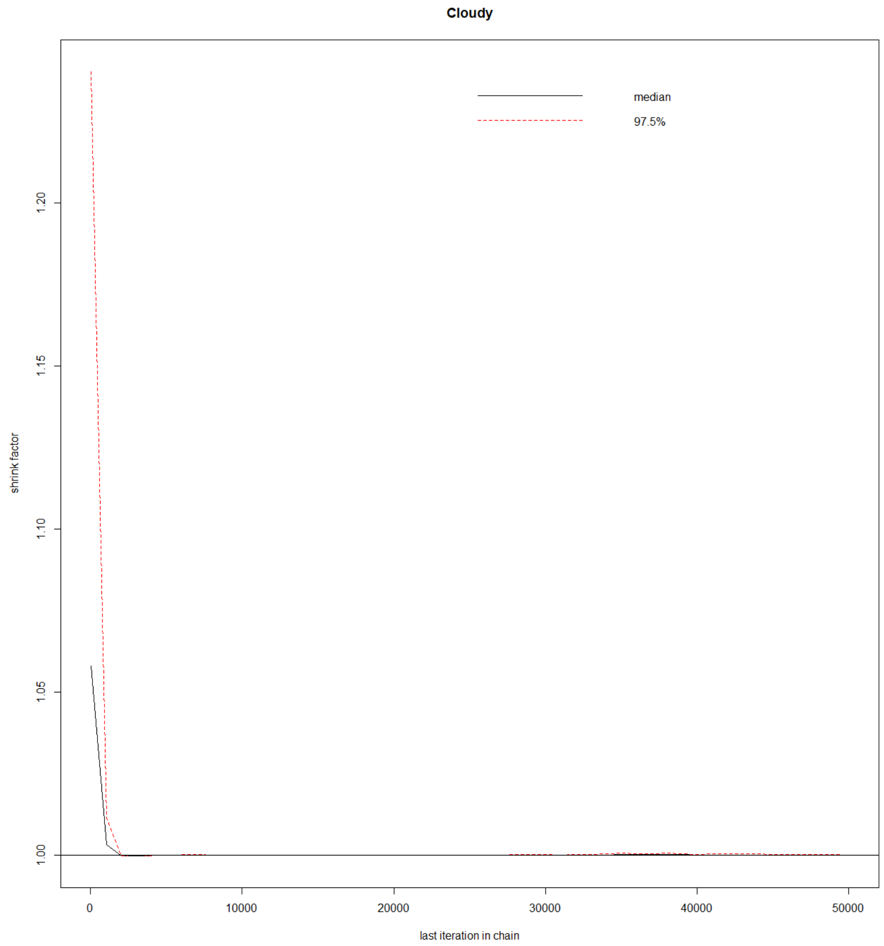
```
The marginal probability of P(R = T | S = T, W = T) calculated by drawing 100 samples from the joint
              probability distribution P(R, C | S = T, W = T) =  0.29
```

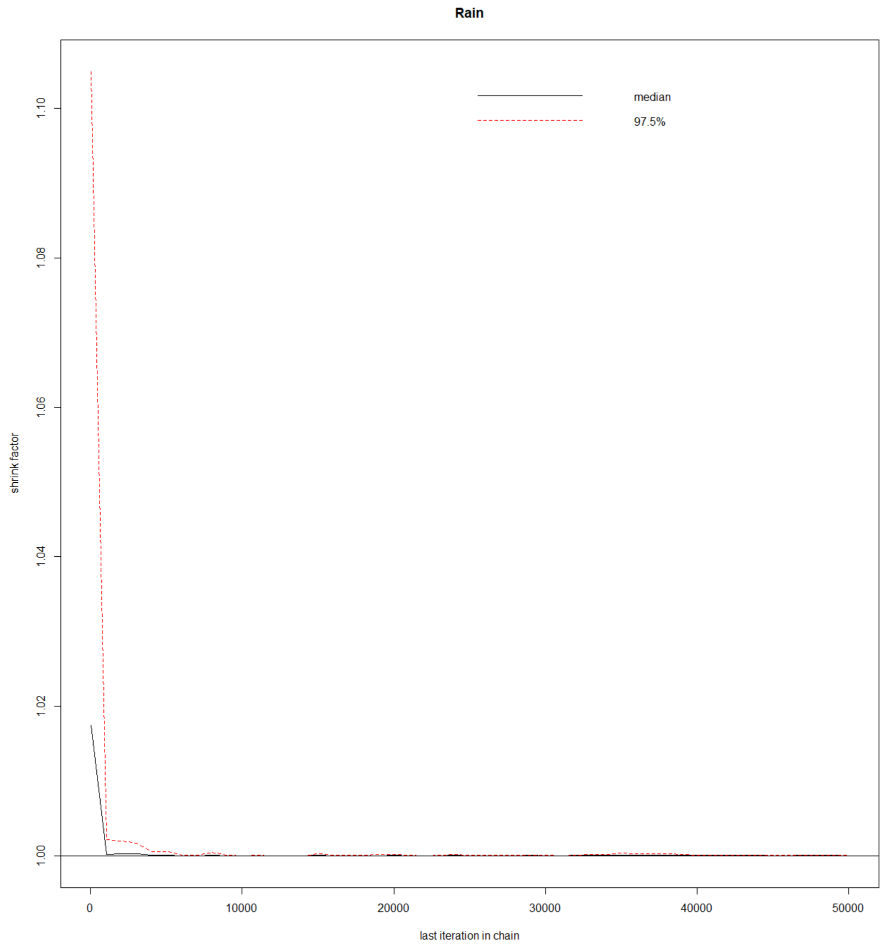
4&5. Now draw 50,000 samples instead of 100 using the Gibbs sampler. Provide the plot of the relative frequencies of R = T and C = T up to each iteration t against t, for two independent runs of the sampler. Suggest a burn-in time based on this plot. (1 point)



Based on the plot above I would suggest a burn-in time value around 5 thousands. (Plots are in high resolution so I'd suggest zooming in to see them better.)

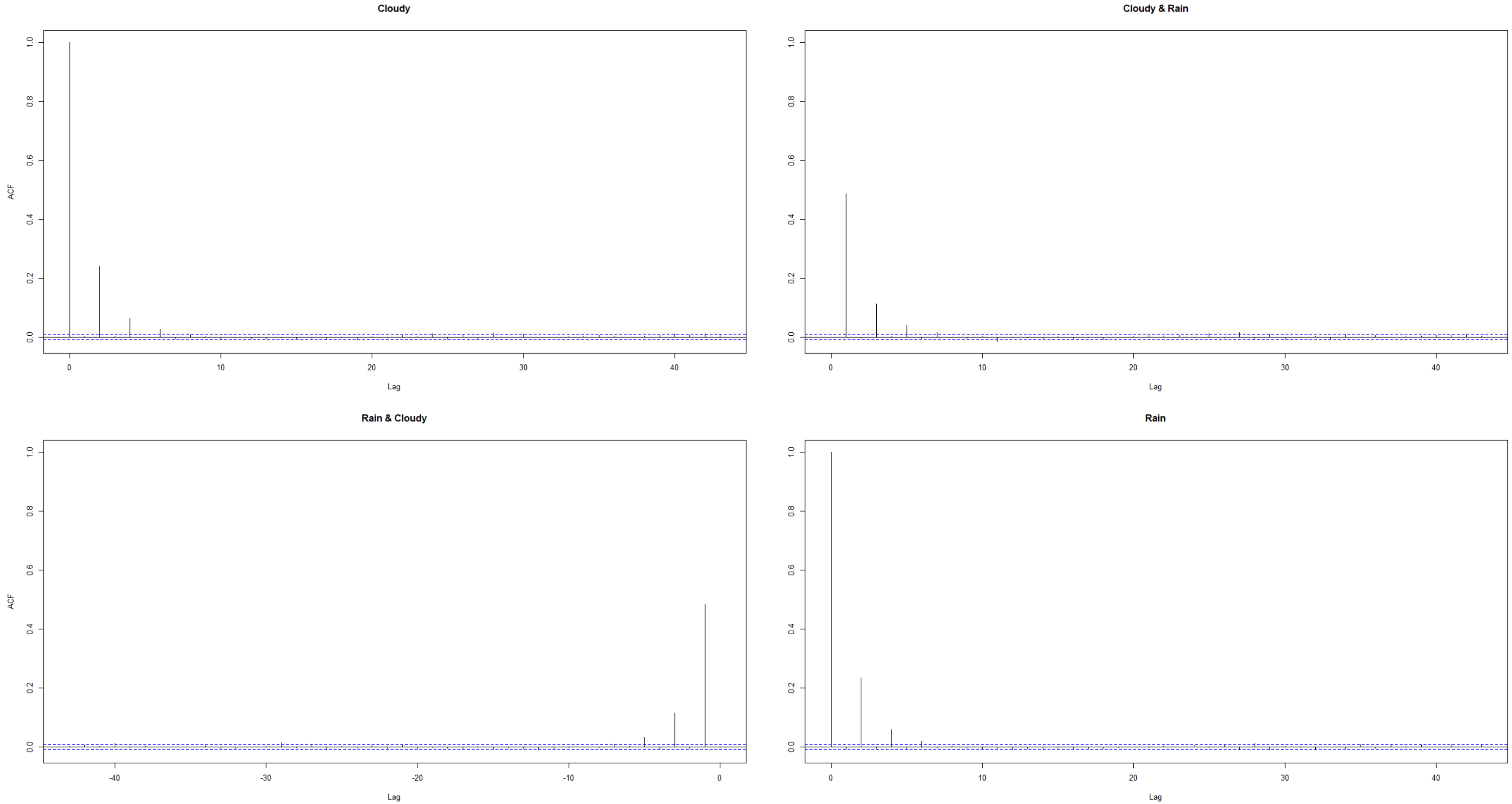
6. Apply the Gelman test and plot potential scale reduction factor changes over the iterations using gelman.plot() from the coda package. Roughly speaking, this factor measures the ratio of the variances within and between independent runs of the sampler. Thus, for a stationary distribution, this factor should be close to 1.0. Suggest a burn-in time based on this plot. (2 points)





Based on this plot, created using gelman.plot() function I would suggest a burn-in time value around 7 thousands.

7. Investigate the auto-correlation among the samples. We expect adjacent members from a Gibbs sampling sequence to be positively correlated, and we can quantify the amount of this correlation by using the auto-correlation function. The lag-k auto-correlation ρ k is the correlation between every draw and its kth neighbouring samples. Use the R-function acf() to provide plots for both variables Rain and Cloudy. Suggest an interval for drawing approximately independent samples. (2 points)



Jugding by the plot above, I would suggest an interval of 4 for drawing approximately independent samples. It seams to by a good compromise between independency and computation time.

8. Re-estimate P (R = T | S = T, W = T) based on 100 samples obtained after the suggested burn-in time and thinning-out. Compare with (3) and comment on your results. (1 point)

The re-estimated probability of P(R = T | S = T, W = T) based on 100 samples obtained after the suggested burn-in (7k) and thinning-out(every 4) = 0.33. Comparing it to the probability value obtained in **3**, which was 0.29, you can see that calculated values aren't that different from one another. The fact that the values are so similar testifies to the robustness of the used estimation method. On the other hand, the results obtained using Gibbs Sampling are dependant on starting point and vary from run to run because of the stochasticity of the sampling process.

9. Compute the probability P(R = T | S = T, W = T) analytically and compare it to the sampling estimate. In real world applications, sampling is performed, because it is usually not possible to easily compute the probabilities analytically. However, since the Bayesian network in Figure 1 is only a small network with discrete variables, the analytical approach is possible. (2 points)

$$P(R = T | S = T, W = T) = \frac{1. = P(R = T, S = T, W = T)}{2. = P(S = T, W = T)}$$

$$1. = P(R = T | C = T) P(S = T | C = T) P(W = T | S = T, R = T) + P(R = T | C = F) P(S = T | C = F) P(W = T | S = T, R = T) = \\ * 0.1 * 0.99 + 0.2 * 0.5 * 0.99 = 0.1782$$

$$2. = P(R = T, S = T, W = T)(= 1.) + P(R = F, S = T, W = T)$$

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

$$P(R = T | S = T, W = T) = \frac{0.1782}{0.1782 + 0.378} = 0.32$$

The probability P(R = T | S = T, W = T) computed analytically is equal to 0.32. The probability P(R = T | S = T, W = T) computed in point **3** = 0.29. The probability P(R = T | S = T, W = T) computed in point **8** = 0.33.

As you can see, in this case the probability obtained in **8** is closer to the probability calculated analytically, than the one from **3**. However, the probability values vary from run to run, so in some cases value from **3** will be closer to the true value, than the value from **8**, but in general values calculated according to scheme from **8** should be closer to the true probability value, than the values calculated according to scheme from **3**.