Assignment 4

Johan Öhman

January 4, 2019

Problem 1

Let X be the numer of colds in a year and let $X \sim Po(\theta)$, where θ is the mean number of colds per year. Normally a person, on average have $\theta = 2$ colds per year, but a new medicine that reduces the number of colds by 50% is tried by the person. So if it works θ is reduced to one. The medicine is known to work for 80% of the population so that

$$P(\theta = 1) = 0.8, \tag{1}$$
$$P(\theta = 2) = 0.2. \tag{2}$$

If the person observes $X = 3$ number of colds in one year, what is then the probability that the medicine worked? What we need is to calculate is the conditional probability $P(\theta = 1|X = 3)$. According to Bayes' theorem this probability can be written as

$$P(\theta = 1|X = 3) = \frac{P(X = 3|\theta = 1)P(\theta = 1)}{P(X = 3)}. \tag{3}$$

The first probability in the numerator is just the pmf at $X = 3$ for the distribution $Po(1)$ which is

$$P(X = 3|\theta = 1) = \frac{e^{-1} \cdot 1^3}{3!} = \frac{1}{6e}, \tag{4}$$

and the second probability in the numerator is given by Eq.(1). Finally the expression in the denominator can be expanded as

$$P(X = 3) = P(X = 3|\theta = 1)P(\theta = 1) + P(X = 3|\theta = 2)P(\theta = 2). \tag{5}$$

Using Eq.(1), Eq.(2) and the pmf for a Poisson distribution, the equation becomes

$$P(X = 3) = \frac{0.8}{6e} + \frac{8 \cdot 0.2}{6e^2}. \tag{6}$$

Inserting the numerical values from equations Eq.(1), Eq.(4) and Eq.(6) into Eq.(3) gives

$$P(\theta = 1|X = 3) \approx 58\%. \tag{7}$$

Problem 2

In the generalized hard-core model the probability of a certain configuration ξ, given the "packing intensity" λ, is as follows

$$\mu_{G,\lambda}(\xi) = \begin{cases} \frac{\lambda^{n(\xi)}}{Z_{G,\lambda}} & \text{if } \xi \text{ is feasible,} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where $n(\xi)$ is the number of ones in that configuration and $Z_{\lambda,G}$ is a normalization constant. The task is to verify the following statement.

As follows from direct calculations, this means that for any vertex $v \in V$, the conditional probability that v takes value 1, given the values at all other vertices equals

$$P(v^{(i)} = 1|v^{(j)} = X_t^j, i \neq j) = \begin{cases} \frac{\lambda}{\lambda+1} & \text{if all neighbours v take value 0,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This is the conditional probability to accept the change at a given vertex and can be viewed as a generalization of the coin-flip in the standard hardcore model. This conditional probability can be written as

$$P(v^{(i)} = 1|v^{(j)} = X_t^j, i \neq j) = \frac{P(v^{(i)} = 1, v^{(j)} = X_t^j, i \neq j)}{P(v^{(j)} = X_t^j, i \neq j).} \tag{10}$$

The probability in the numerator is the probability that a certain configuration ξ_1 occurs and the probability in the denominator is the sum of probabilities for the configurations ξ_1 and ξ_2 given as

$$\begin{aligned} P(v^{(j)} = X_t^j, i \neq j) &= P(\xi_1) + P(\xi_2) && (11) \\ &= P(v^{(i)} = 1|v^{(j)} = X_t^j, i \neq j) + P(v^{(i)} = 0|v^{(j)} = X_t^j, i \neq j) && (12) \\ &= \frac{\lambda^{n(\xi_1)}}{Z_{G,\lambda}} + \frac{\lambda^{n(\xi_2)}}{Z_{G,\lambda}}. && (13) \end{aligned}$$

The conditional probability can then be written as

$$P(v^{(i)} = 1|v^{(j)} = X_t^j, i \neq j) = \frac{\frac{\lambda^{n(\xi_1)}}{Z_{G,\lambda}}}{\frac{\lambda^{n(\xi_1)}}{Z_{G,\lambda}} + \frac{\lambda^{n(\xi_2)}}{Z_{G,\lambda}}} = \frac{\lambda^{n(\xi_1)}}{\lambda^{n(\xi_1)} + \lambda^{n(\xi_2)}}, \tag{14}$$

since that the only difference between the configurations is that the vertex $x^{(i)} = 1$ for ξ_1 and $x^{(i)} = 0$ for ξ_2 hence $n(\xi_1) = n(\xi_2) + 1$ and the probability can be rewritten as

$$P(v^{(i)} = 1|v^{(j)} = X_t^j, i \neq j) = \frac{\lambda^{n(\xi_2)+1}}{\lambda^{n(\xi_2)+1} + \lambda^{n(\xi_2)}} = \frac{\lambda}{\lambda + 1}, \tag{15}$$

and thus the statement is verified.

Problem 3

A stationary distribution π fulfills

$$\pi = \pi \mathbf{P}, \tag{16}$$

but it should also fullfill

$$\pi = \pi P^n, \quad n \to \infty, \tag{17}$$

this can be written on component form as

$$\pi_j = [\pi \mathbf{P}^n]_j = \sum_{i=1}^{k} \pi_i [\mathbf{P}^n]_{i,j} = \sum_{i=1}^{k} P(X_0 = s_i) \; P(X_n = s_j|X_0 = s_i) \tag{18}$$

If the state s_j is non-essential so that s_j is not part of a communicating class then the probability to reach that state as $n \to \infty$is zero, i.a

$$P(X_0 = s_i)\ P(X_n = s_j | X_0 = s_i) = 0, \quad \forall\ i, \quad \text{as } n \to \infty, \tag{19}$$

inserting this into Eq.(18) it becomes clear that the stationary distribution does not give any mass for a nonessential state s_j.

Problem 4

X is a convex subset of $\mathbb{R}^k$. The stationary distribution $\pi \in X$ is ergodic, which means that it put all its mass on a single communicating class and that the stationary distribution is unique on that communicating class. The task is to show that π is an extreme point on X. A point on a convex set can, in general, be written as

$$\pi = p\pi'' + (1-p)\pi', \tag{20}$$

where π' and π'' are some other points in the set and $p \in [0, 1]$ is a constant. Since all terms in Eq.(20) are positive the only way that π can be written on this form is if π' and π'' also put all its mass on the same communicating class as π. But it is known that π is unique on that communicating class so the only solution that exists is

$$\pi' = \pi'' = \pi, \tag{21}$$

which makes it impossible to write π as a linear combination as in Eq.(20), hence π is an extreme point on X.

Problem 5

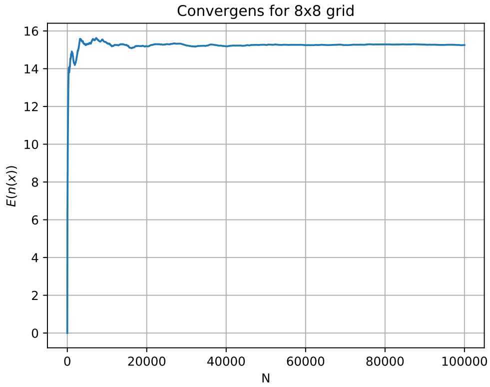
a) The time averaged number of ones in the hard-core model $\frac{1}{N}\sum_{t=1}^{N-1} n(X_t)$ is close to the expected number of ones

$$E(n(X)) = \frac{1}{Z} \sum_{\xi \text{ feasible}} n(\xi), \tag{22}$$

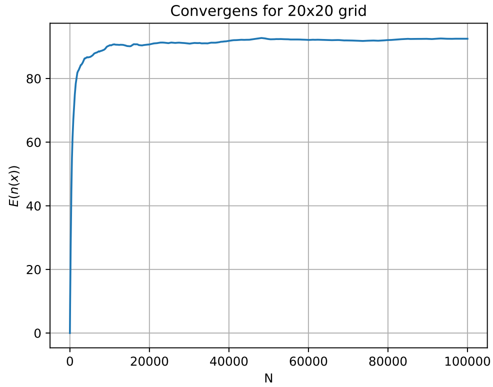
since the Markov chain X_t for large N will have transited to all the different configurations possible and have the stationary distribution equal to the possibilities to choosing a certain configuration.
b)
The expected number of ones $E(n(x))$ is to be estimated using a MCMC simulation of the hard-core model. Starting with an empty grid the following algorithm is used to estimate $E(n(X))$

1. Pick a point on the grid randomly with uniform probability.
2. Check if horizontal and vertical neighbours are all zero. If so proceed, otherwise start over from step 1.
3. Accept the point with 50% probability and then change the value of that vertex.
4. Compute $\frac{1}{N}\sum_{t=1}^{N-1} n(X_t)$
5. Repeat from step 1.

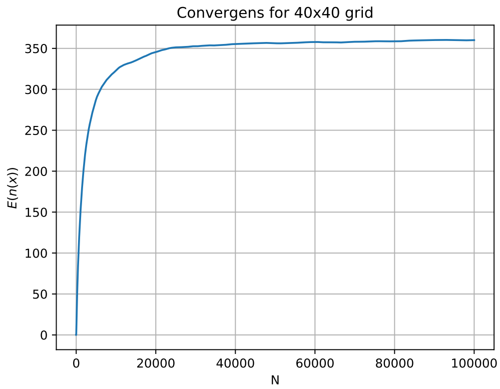
In Figure 1 the estimate of $E(n(X))$ is shown as function of N iterations for (a) a 8x8 grid (b) a 20x20 grid and (c) a 40x40 grid. For the 8x8 grid $E(n(X)) \approx 15$, for the 20x20grid $E(n(X)) \approx 95$ and for the 40x40 $E(n(X)) \approx 360$. This seams reasonable since, for example, for the 8x8 grid the maximum number of ones would be 32 (every other square is a one) and one would expect around half of that value.



(a)



(b)



(c)

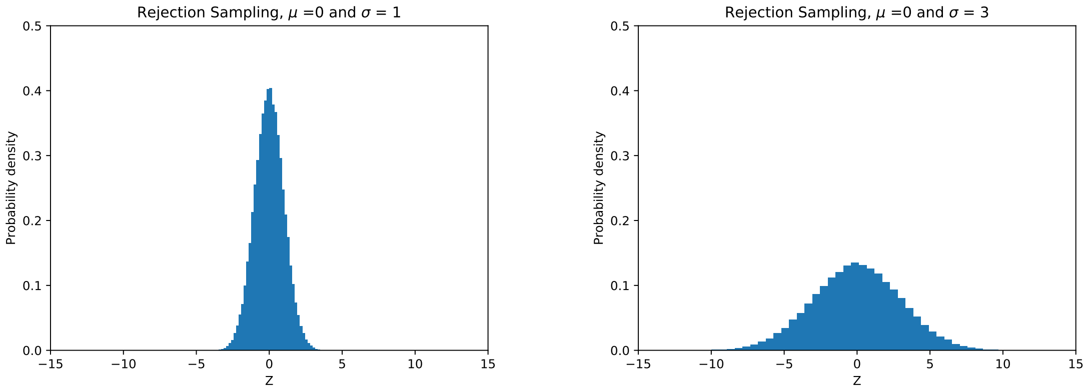
Figure 1: Convergence of the estimate of $E(n(X))$ for (a) a 8x8 grid (b) a 20x20 grid and (c) a 40x40 grid.

Problem 6

We want to sample a random variable from the target distribution $Z \sim N(\mu, \sigma)$ using rejection sampling. Let $\varphi(x)$ be the pdf of the $N(0,1)$ distribution and let X Exp(1) be a sample from the proposal distribution with pdf $f(x)$. X can be sampled as $X = -\log(rand())$ using the inverse transform method. To sample from $Z \sim N(\mu, \sigma)$ the following algorithm is applied

1. Sample $X \sim Exp(1)$
2. Sample $Y \sim U(0,1)$
3. Check if $\frac{\varphi(X)}{c \cdot f(X)} > Y$, where c is a constant. If so proceed, otherwise reject the sample.
4. Sample $Y_2 \sim U(0,1)$
5. If $Y_2 > 0.5$ save $Z = X \cdot \sigma + \mu$, otherwise save $Z = -X \cdot \sigma + \mu$.

Step 5 is needed since the exponential distribution only is defined for positive numbers. But since the normal distribution is symmetric around μ the sign can be set in this way. To verify that the rejection sampling works the sampling is repeated 100000 times for both $N(0,1)$ and $N(0,3)$. The result is shown in Figure 2 and it looks that the rejection sampling works as intended.



(a) (b)

Figure 2: Histogram obtained using rejection sampling from (a) a $N(0,1)$ distribution and (b) a $N(0,3)$ distribution.

Problem 7

In this task the Metropolis-Hastings algorithm will be used to estimate the kurtosis $E(X^4)$ of the t-distribution with pdf

$$f(x) = c(1 + \frac{x^2}{d})^{-(d+1)/2}, -\infty < x < \infty, \tag{23}$$

where c is a constant and in this case $d = 5$. Let X be a random variable from the t-distribution. To sample X we use a MCMC simulation. A proposed sample x' is sampled from the proposal distribution $q(x'|X_t)$. Given x' and the current state X_t the acceptance probability is

$$\alpha = \min\left(1, \frac{f(x')}{f(X_t)} \cdot \frac{q(X_t|x')}{q(x'|X_t)}\right). \tag{24}$$

The proposed x' is accepted with the probability α. If accepted the next state in the chain becomes $X_{t+1} = x'$ otherwise $X_{t+1} = X_t$

In practice the algorithm can be summarized as

1. Initiate the chain with $X_0 = 0$.
2. Then at a time t: sample $x' \sim N(X_t, \sigma)$ using the rejection sampling algorithm.
3. Compute the acceptance probability α.
4. Sample $Y \sim U(0,1)$.
5. If $U < a$ set $X_{t+1} = x'$, otherwise $X_{t+1} = X_t$.
6. Repeat from step 2.

Two different proposal distributions are used in the solution to this problem, the first one is the basis for the so called Metropolis random walk algorithm where the proposal distribution is the pdf of $N(X_t, \sigma)$

$$q(x'|X_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left((x' - X_t)^2/2\sigma\right), \tag{25}$$

note that $q(x'|X_t) = q(X_t|x')$ so that α reduces to

$$\alpha = \min\left(1, \frac{f(x')}{f(X_t)}\right). \tag{26}$$

The Metropolis random walk algorithm is used with $\sigma = [1, 2, 5]$ and $N = 100000$ time steps. After each time step the estimate of $E(X^4)$ is computed. In Figure 3 the histograms of X along with the evaluation of $E(X^4)$ over time for the different standard deviations are showed. It seams that it converges towards $E(X^4) \approx 13$ for $\sigma = [1, 2]$ while when $\sigma = 5$ the convergence is slower an no final value have been reached at the final time step.

The second proposal distribution is the pdf of a $N(0, \sigma)$ distribution

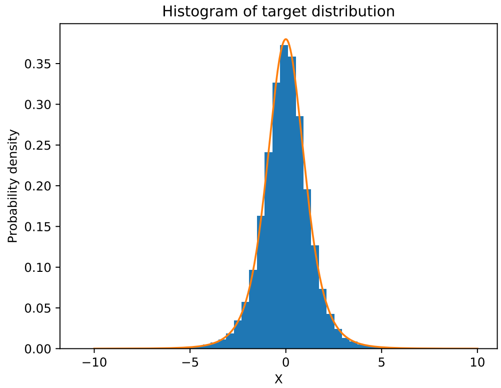
$$q(x'|X_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left((x')^2/2\sigma\right), \tag{27}$$

it can be noted that the proposed values x' will be independent from the current state X_t, this was not the case using the first proposal distribution. Again the Metropolis-Hastings algorithm is used with $\sigma = [1, 7, 10]$ and $N = 100000$ time steps and after each step $E(X^4)$ is computed. In Figure 4 the histograms of X along with the evaluation of $E(X^4)$ over time for the different standard deviations are showed. Looking at the first row where $\sigma = 1$ it is clear that X do not have the same distribution as the target distribution. It is required that $q(x|X_t) > f(x)\ \forall x$ for the algorithm to work, and that is not the case here. For $\sigma = [7, 10]$ the distribution of X converges to the target and $E(X^4) \approx 13$. Using this proposal distribution it seams that the convergence is quicker which should be no surprise since we added information that $\mu = 0$ for the target distribution.

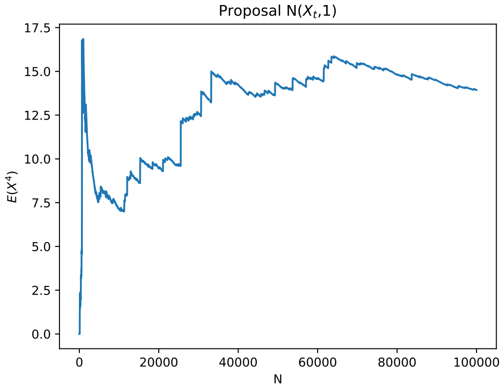
Overall all the cases might have needed to run for additional time-steps, however, my implementation became very slow for $N > 100000$ so I settled for theses time steps. Since the kurtosis is to the power four large value in the chain can offset the convergence significantly.

Code

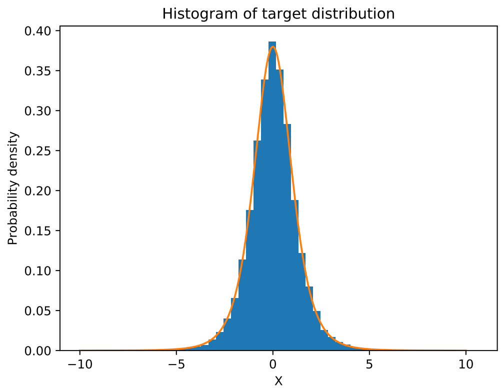
All simulations have been made using Python 3 and the numpy package. The scripts used are available at my Github Repository `https://github.com/hyllevask/MCMC_simulations`



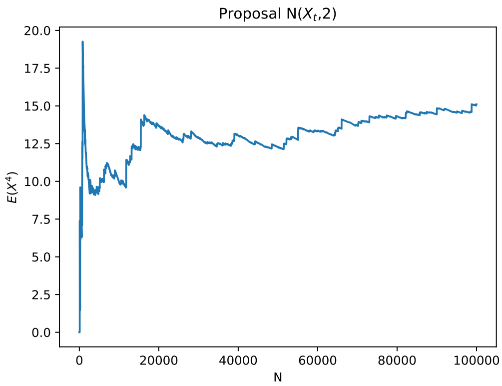
(a)



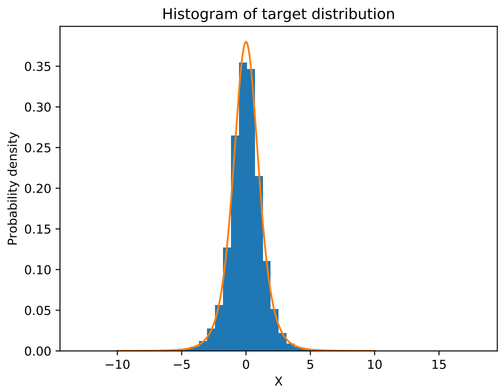
(b)



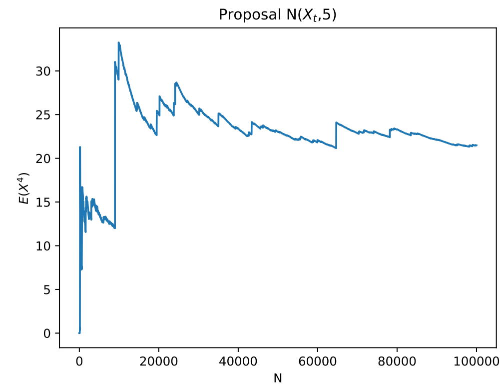
(c)



(d)

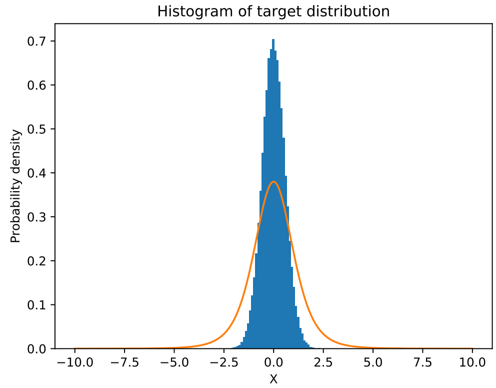


(e)

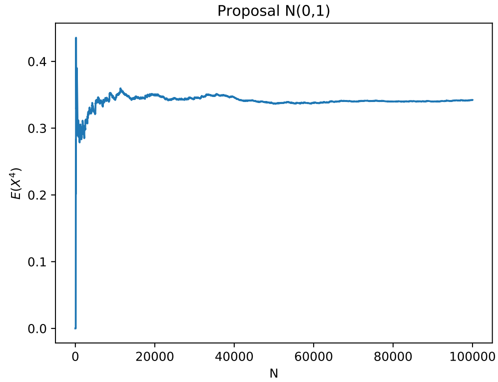


(f)

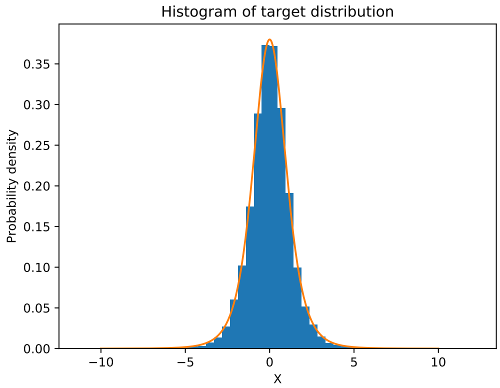
Figure 3: MCMC simulation of the t-distribution using the Metropolitan random walk algorithm. (a),(b),(c) shows the histogram of X and the pdf of the t-distribution and (b),(d),(f) are the corresponding MCMC estimates of the kurtosis $E(X^4)$. (a) and (b) uses the proposal distribution $N(X_t, 1)$, (c) and (d) uses $N(X_t, 2)$ and (e) and (f) uses $N(X_t, 5)$



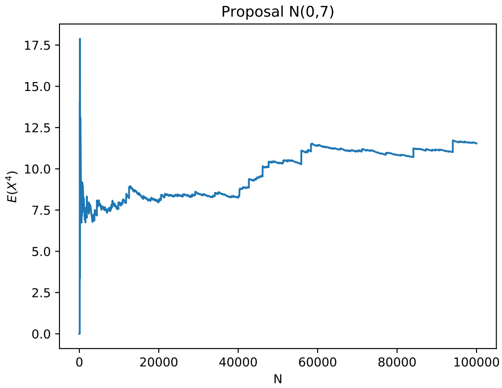
(a)



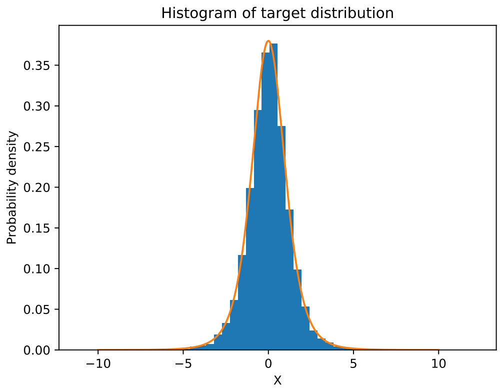
(b)



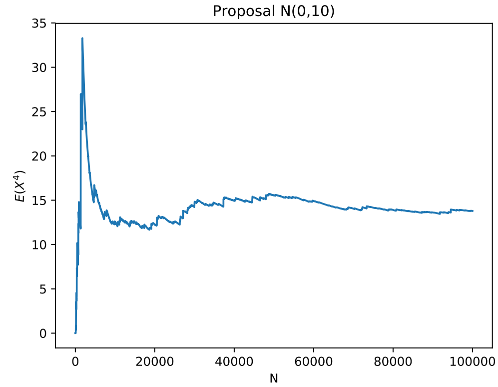
(c)



(d)



(e)



(f)

Figure 4: MCMC simulation of the t-distribution using the Metropolitan-Hastings algorithm. (a),(b),(c) shows the histogram of X and the pdf of the t-distribution and (b),(d),(f) are the corresponding MCMC estimates of the kurtosis $E(X^4)$. (a) and (b) uses the proposal distribution $N(0,1)$, (c) and (d) uses $N(0,7)$ and (e) and (f) uses $N(0,10)$