Stochastic Simulations
Homework 3: Markov chain Monte Carlo

Bruno Rodriguez Carrillo
EPFL

December 9, 2020

1. To compute the expected acceptance probability for the independent Metropolis-Hastings algorithm, we proceed as follows:

   In case of the independence sampler, we have the following expression for acceptance probability $\alpha(\boldsymbol{x}, \boldsymbol{y})$:

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{\frac{f(\boldsymbol{y})g(\boldsymbol{x})}{f(\boldsymbol{x})g(\boldsymbol{y})}, 1\right\}$$

   As a result, expected acceptance probability can be computed as:

$$\mathbb{E}\left(\alpha(\boldsymbol{x}, \boldsymbol{y})\right) = \mathbb{E}\left(\min\left\{\frac{f(\boldsymbol{y})g(\boldsymbol{x})}{f(\boldsymbol{x})g(\boldsymbol{y})}, 1\right\}\right)$$

   Let us $l(\boldsymbol{y}, \boldsymbol{x}) = \dfrac{f(\boldsymbol{y})g(\boldsymbol{x})}{f(\boldsymbol{x})g(\boldsymbol{y})}$. As a result:

$$\mathbb{E}\left(\min\left\{\frac{f(\boldsymbol{y})g(\boldsymbol{x})}{f(\boldsymbol{x})g(\boldsymbol{y})}, 1\right\}\right) = \iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})\geq 1} f(\boldsymbol{x})g(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} + \iint_{\mathcal{X}} l(\boldsymbol{y}, \boldsymbol{x})\mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})<1} f(\boldsymbol{x})g(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}$$

$$= \iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})\geq 1} f(\boldsymbol{x})g(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} + \iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})<1} f(\boldsymbol{y})g(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}$$

$$= 2\iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})\geq 1} f(\boldsymbol{x})g(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \geq 2\iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})\geq 1} f(\boldsymbol{x})f(\boldsymbol{y})\frac{1}{C}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} =$$

$$= \frac{2}{C}\iint_{\mathcal{X}} \mathbb{1}_{l(\boldsymbol{y},\boldsymbol{x})\geq 1} f(\boldsymbol{y})f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} = \frac{1}{C}$$

   The inequality comes from the fact that there exists a constant C, such that $\dfrac{1}{C}f(\boldsymbol{x}) \leq g(\boldsymbol{x})$.

   This means that $\mathbb{E}\left(\min\left\{\dfrac{f(\boldsymbol{y})g(\boldsymbol{x})}{f(\boldsymbol{x})g(\boldsymbol{y})}, 1\right\}\right) \geq \dfrac{1}{C}$.

   On the other hand, the acceptance rate for the acceptance-rejection method is always $\dfrac{1}{C}$.

2. The goal is to sample from a Gamma distribution with shape parameter α and scale parameter β, denoted by Gamma(α,β), so that the target PDF reads:

$$f(x) = f(x;\alpha,\beta) = \beta^{\alpha}x^{\alpha-1}\mathrm{e}^{\beta x}/\Gamma(\alpha)\mathbb{1}_{x\geq 0} \tag{1}$$

   where $\Gamma(\cdot)$ denotes the Gamma function.

   All simulations hereafter are performed with $\alpha = 4.85$. Besides, $n \in \{250, ..., 5000\}$, which means n takes 20 different values.

   (a) Implement the Acceptance-Reject method to sample from Gamma(α, 1) for $\alpha \geq 1$, using the PDF of the Gamma(α,β) distribution with $a = [\alpha]$ as auxiliary density (here $[\alpha]$ denotes the integer part of α). Show that $b = \alpha/[\alpha]$ is the optimal choice for b.

   To sample from the Gamma(α, β) distribution, we precede as follows:

First, we generate a $\text{Gamma}(1,\beta) = \text{Exp}(\beta)$ distribution. Then we generate $\sum_{j=1}^{K} \xi_j \sim \text{Gamma}(K,\beta)$. Taking into account the values of our parameters, we obtain:
$\sum_{j=1}^{[\alpha]} \xi_j \sim \text{Gamma}([\alpha],[\alpha]/\alpha)$
And $\xi_j \sim \text{Gamma}(1,[\alpha]/\alpha) = \text{Exp}([\alpha]/\alpha)$
As a reminder, the efficiency of the acceptance-rejection method is defined as the probability of acceptance:

$$P\left(U \leq \frac{f(X)}{Cg(X)}\right) = \frac{1}{C} \tag{2}$$

In our case, $f(X)$ and $g(X)$ in (2) are defined in (1) and $X \sim \text{Gamma}([\alpha],[\alpha]/\alpha)$.
Consequently, the value of C in the AR method is computed as follows:

$$C \geq \frac{f(x;\alpha,1)}{f(x;[\alpha],b)} \Rightarrow C \geq \max\left\{\frac{x^{\alpha-[\alpha]}\mathrm{e}^{x(b-1)b^{[-\alpha]}}\Gamma([\alpha])}{\Gamma(\alpha)}\right\}$$

Let $h(x) = \dfrac{f(x;\alpha,1)}{f(x;[\alpha],b)}$, then to obtain the maximum value of $h(x)$, we compute its derivative with respect to x and equal it to 0, that is:
$\dfrac{\mathrm{d}}{\mathrm{d}x}h(x) = 0 \Rightarrow \dfrac{b^{[\alpha]}x^{[\alpha]-\alpha}\left([\alpha]-\alpha\right)\mathrm{e}^{-x(b-1)}\Gamma\left(\alpha\right)}{x\Gamma\left([\alpha]\right)} + \dfrac{b^{[\alpha]}x^{[\alpha]-\alpha}\left(-b+1\right)\mathrm{e}^{-x(b-1)}\Gamma\left(\alpha\right)}{\Gamma\left([\alpha]\right)} = 0$
Solving for x, we obtain the critical value of x which maximizes $h(x)$, which is: $x = \dfrac{[\alpha]-\alpha}{b-1}$.
Now, in order to know the optimal value for b, we substitute such a value of x in $h(x)$, calling $k(b)$ the result of such a substitution, we have:

$$k(b) = \frac{\mathrm{e}^{[\alpha]-\alpha}b^{-[\alpha]}\Gamma\left([\alpha]\right)}{\Gamma\left(\alpha\right)}\left(\frac{[\alpha]-\alpha}{b-1}\right)^{\alpha-[\alpha]}$$

Then, we take the derivative $\dfrac{\mathrm{d}}{\mathrm{d}b}k(b)$ and equal it to 0, that is $\dfrac{\mathrm{d}}{\mathrm{d}b}k(b) = 0$, which gives:

$$\frac{\mathrm{d}}{\mathrm{d}b}k(b) = \frac{\mathrm{e}^{[\alpha]-\alpha}b^{-[\alpha]-1}\left(-b\alpha+[\alpha]\right)\Gamma\left([\alpha]\right)}{\left(b-1\right)\Gamma\left(\alpha\right)}\left(\frac{[\alpha]-\alpha}{b-1}\right)^{\alpha-[\alpha]} = 0$$

Solving for b, we have the optimal value for b: $b = \dfrac{[\alpha]}{\alpha}$.
Consequently, we have the following expression for C in AR method:

$$C \geq (\alpha)^{\alpha-[\alpha]}\left(\mathrm{e}\right)^{[\alpha]-\alpha}\left(\frac{[\alpha]}{\alpha}\right)^{-[\alpha]}\frac{\Gamma([\alpha])}{\Gamma(\alpha)} \tag{3}$$

Then, taking $\alpha = 4.85$, we obtain the minimum value for C, that is: $C = 1.105143$. This is the optimal choice for C. The closest C is to its optimal value, the larger the acceptance is for AR method.
As consequence, implementing the algorithms, we obtained the following results for both methods:

method	mean estimator	acceptance rate
AR	4.857	0.908
MH	4.861	0.934

Table 1: Result comparison for $n = 5000$.

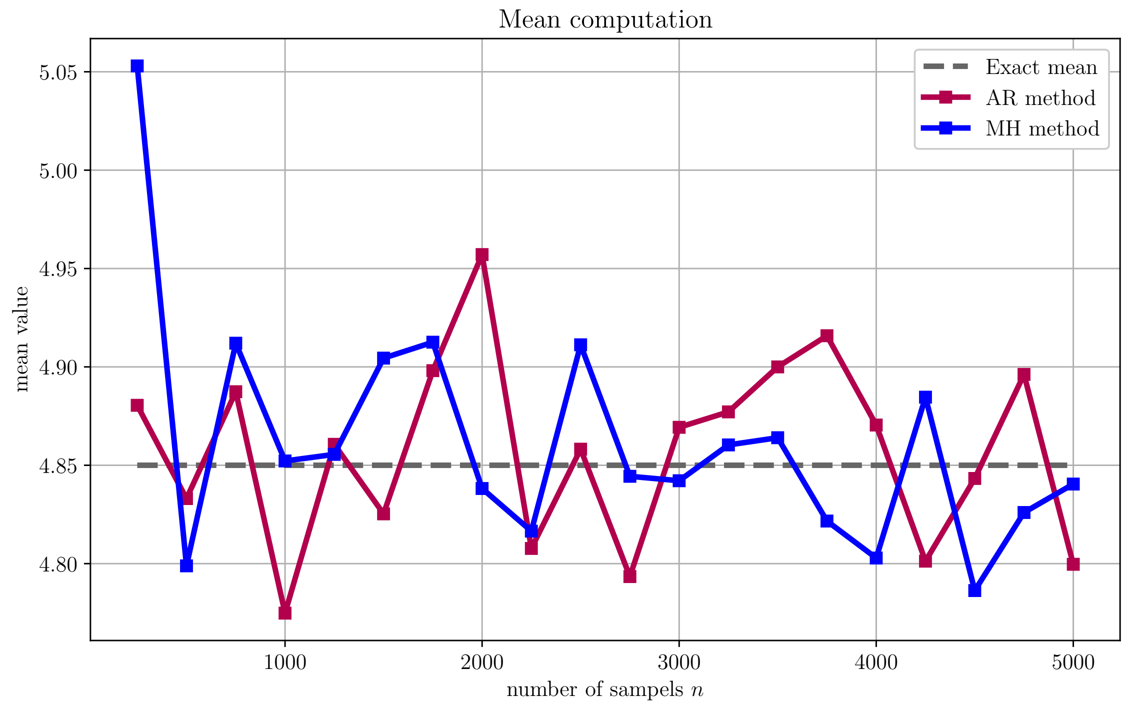


Figure 1: Mean computation comparison.

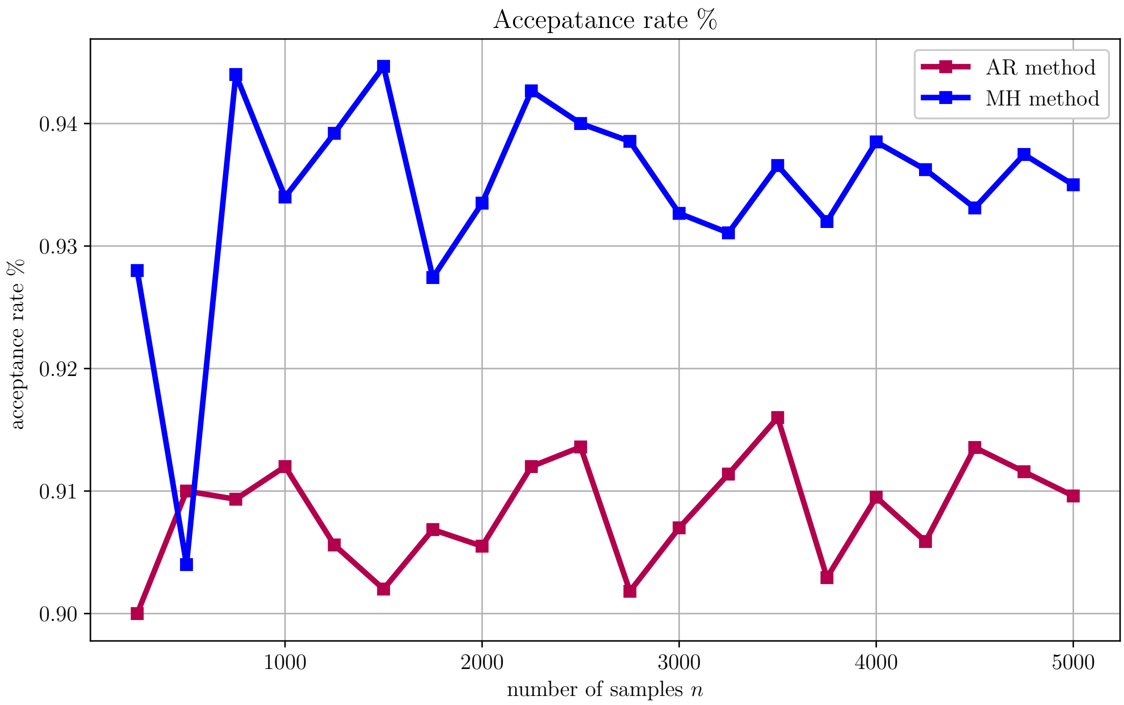
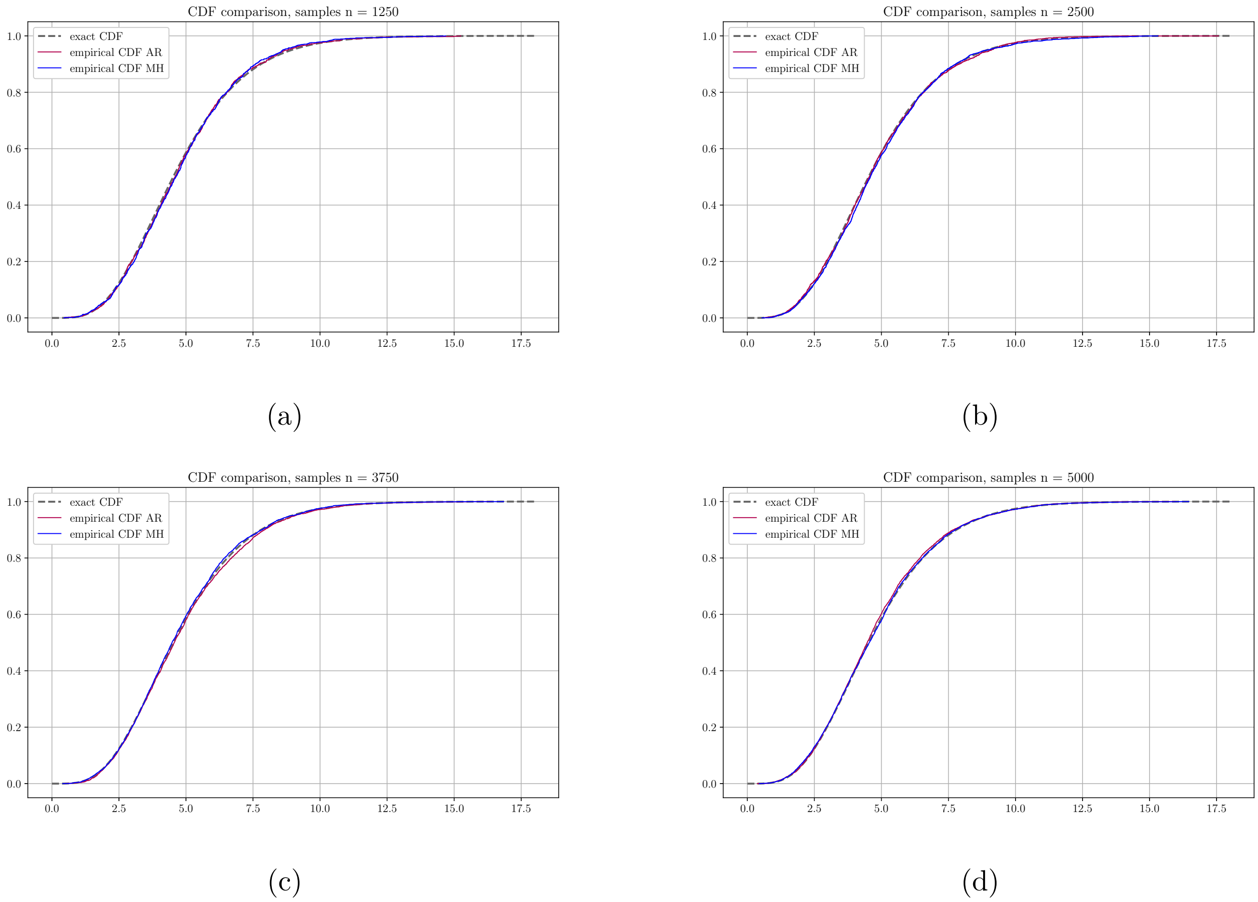


Figure 2: Acceptance rate comparison.



(a) (b) (c) (d)

Figure 3: CDF comparison.

From these plots, we can observe that both AR and MH algorithms display quite similar results regarding the estimate for the mean value of the target distribution $f(x; \alpha, 1)$, which is $Gamma(x; 4.85, 1)$.

Besides, according to chart (1), the acceptance rate for MH is higher than the one computed via AR. This corresponds to what we showed before. In order words, the acceptance rate for MH is at least equaled to the acceptance rate for AR. In our case $C_{MH} > C_{AR}$. However, although this holds, there is no a significant difference between the computed mean values since both approximated the mean of the target distribution accurately.

It also worth noting that the value of C_{AR} was chosen as the optimal and for such a value, the acceptance rate for AR is the highest.

As expected, the higher the number of samples is, the more accurate our results are, which can be observed in figures (3). We can conclude that AR and MH methods performed in such a manner that the final distribution is the one we were expecting.