This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3038136, IEEE Signal Processing Letters



# An insight into the Gibbs sampler: keep the samples or drop them?

Julien Boissy, Jean-François Giovannelli and Pierre Minvielle

***Abstract*—In this letter, we propose an insight into Markov Chain Monte Carlo (MCMC) algorithms and more precisely the Gibbs sampler. From a didactic toy model, based on a normal bivariate distribution, a non-asymptotic analysis is derived and estimators are fully characterized. It provides a worthwhile and non-empirical understanding of the Gibbs sampler behaviour. Issues are investigated, such as the influence of the "burn-in" phase, useful in practice. Especially, the trade-off between discarding samples and integrating them into estimators is studied. On the whole, it leads to an analytical awareness of MCMC sampler.**

***Index Terms*—MCMC, Bayesian statistics, Gibbs, burn-in.**

## I. Introduction

THE Gibbs sampler is an ubiquitous MCMC algorithm in Bayesian statistics and machine learning [1], [2]; it was introduced in the context of image processing by [3]. Especially applied to hierarchical models, the Gibbs sampler is, as well as the Metropolis-Hastings (MH) sampler of which it is a special case, the workhorse of general MCMC techniques. Hereafter, we provide an analytical insight into the standard Gibbs algorithm and its mechanisms.

We consider a couple of real random variables $(X, Y)$ of joint probability density function $f_{X,Y}$. Let us suppose that though it is difficult to sample from it directly, realisations can easily be obtained from the conditional distributions $f_{X|Y}$ and $f_{Y|X}$. The elementary two-stage Gibbs sampler, see below, is appropriate in such a context [1]. It generates a realisation of the Markov chain $(X^{(n)}, Y^{(n)})_{n\geq 1}$ that has the target distribution $f_{X,Y}$ as invariant distribution. It is typically used to solve intractable integration problems [2], central to Bayesian statistics, such as expectation $E_{f_{XY}}[h(X,Y)] = \iint h(x,y) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ or marginalization $f_X(x) = \int f_{X,Y}(x,y)\,\mathrm{d}y$.

---

**Algorithm** Two-stage Gibbs sampler

---

**Input:** $N, N_0, f_{Y^{(0)}}$
  $y^{(0)} \sim f_{Y^{(0)}}(\cdot)$
  **for** $n = 1 \cdots N$ **do**
    $x^{(n)} \sim f_{X|Y}(\cdot|y^{(n-1)})$
    $y^{(n)} \sim f_{Y|X}(\cdot|x^{(n)})$
  **end for**
**Output:** samples $(x^{(n)}, y^{(n)})_{N_0 \leq n \leq N}$

---

The algorithm is parametrized by $N$, the chain run length and $N_0$, the so-called "burn-in" period during which samples are assumed not to be samples of the target distribution and discarded in order to reduce starting biases [2]. The initial value of $Y$ is randomly drawn from the density $f_{Y^{(0)}}$. At the end, the Gibbs output is made of the samples: $(x^{(n)}, y^{(n)})_{N_0 \leq n \leq N}$.

The determination of the appropriate lengths $N$ and $N_0$ is known to be a tricky task. How long must the "burn-in" period $N_0$ be chosen? How many samples $\Delta N \triangleq N - N_0 + 1$ are required to accurately approximate for instance $E_{f_{XY}}[h(X,Y)]$? For a fixed $N$, the higher $N_0$ is, the "better" the samples are, but the fewer there are. What is the optimal trade-off between the sample length and the sample quality? Conversely, for a fixed $N_0$, the higher $N$ is, the more accurate the approximation is, but the longer the Gibbs sampler is.

Similarly to the seminal work by Gelman et al. [4] for the Metropolis-Hastings algorithm, its optimal scale factor and its known acceptance rate of $0.234$, we intend to examine critical parameters of the Gibbs sampler, such as the "burn-in" length. As extensively done in [4], our exploration makes use of a toy model; it is of great interest to investigate such intricate issues. Although it can be related to basic examples in [5, p. 131], [1, p. 340] and [6], we do believe that we provide an attractive investigation into MCMC algorithms and the "burn-in" issue. We consider for instance the following straightforward MCMC estimate or approximation of $E[Y]$, e.g. the posterior mean in the Bayesian setting; it consists of the empirical mean of the samples $(y^{(n)})_{N_0 \leq n \leq N}$ from the Gibbs output:

$$E[Y] \approx \frac{1}{\Delta N} \sum_{n=N_0}^{N} y^{(n)}. \tag{1}$$

Our main contribution is therefore to develop closed-form expressions for the bias and the variance of related estimators, from which efficient and meaningful analyses can be performed. It leads to a new study of the influence of the "burn-in" length $N_0$. It differs from former works about Gibbs sampling: [7] is based on subsampling and stationarity considerations, [8]–[11] propose improvements where numerical experiments are developed for performance assessment and [12] uses bivariate Gaussian toy examples again for numerical experiments. Even more recent works, such as [12]–[15], do not develop such a theoretical understanding of "burn-in".

## II. Normal bivariate toymodel

We focus on a Gaussian bivariate toy model: $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu} = [\mu_X\ \mu_Y]^T$ and covariance matrix

J. Boissy and P. Minvielle are with CESTA, DAM, CEA, F-33114, Le Barp, France, E-mail: julien.boissy@cea.fr; pierre.minvielle@cea.fr. J.-F. Giovannelli is with Groupe Signal-Image, IMS (Univ. Bordeaux, CNRS, BINP), F-33400 Talence, France, E-mail: Giova@IMS-Bordeaux.fr.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3038136, IEEE Signal Processing Letters



$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho \cdot \sigma_X \sigma_Y \\ \rho \cdot \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}. \quad (2)$$

Hereinafter, $f_{X,Y}$ represents the target distribution. It is well-known that the conditional laws are normal (see [16, p. 43-44]):

$$X|Y = y \sim \mathcal{N}(\mu_{X|y}, \sigma^2_{X|y}), \quad (3)$$

where $\mu_{X|y} = \mu_X + \rho\sigma_X(y - \mu_Y)/\sigma_Y$ and $\sigma^2_{X|y} = \sigma^2_X(1 - \rho^2)$, and symmetrically for $Y|X = x$. In addition, the initial distribution is defined by:

$$f_{Y^{(0)}}(y^{(0)}) = \mathcal{N}(y^{(0)}; m_{Y^{(0)}}, \sigma^2_{Y^{(0)}}). \quad (4)$$

The algorithm determines a realisation of the Markov chain: $y^{(0)} \to x^{(1)} \to y^{(1)} \to x^{(2)} \to y^{(2)} \to \cdots \to x^{(N)} \to y^{(N)}$, leading to the successive couples $(x^{(n)}, y^{(n)})_{1\le n\le N}$. It is straightforward to check that $f_{X,Y}$ is a stationary distribution [1]. Next, we determine analytical expressions of the marginal probability densities $f_{X^{(n)}}$ and $f_{Y^{(n)}}$. After, we focus on the MCMC estimator of the expectation, such as $E(Y)$.

### A. Elementary AR(1) model

Let us note $\mathcal{X}^{(n)} \triangleq (X^{(n)} - \mu_X)/\sigma_X$ and $\mathcal{Y}^{(n)} \triangleq (Y^{(n)} - \mu_Y)/\sigma_Y$, it is direct to exhibit an elementary auto-regressive $AR(1)$ model (see toy models [1, p. 340] and [17, p. 9]):

$$\mathcal{Y}^{(n+1)}|\mathcal{Y}^{(n)} = y^{(n)} \sim \mathcal{N}(\rho^2 y^{(n)}, 1 - \rho^4). \quad (5)$$

Noting the expectation as $m_{\mathcal{Y}^{(n)}} \triangleq E[\mathcal{Y}^{(n)}]$ and the variance as $\sigma^2_{\mathcal{Y}^{(n)}} \triangleq V(\mathcal{Y}^{(n)})$, the recursive expressions are obtained:

$$m_{\mathcal{Y}^{(n+1)}} = \rho^2 m_{\mathcal{Y}^{(n)}}, \quad \sigma^2_{\mathcal{Y}^{(n+1)}} = 1 + \rho^4(\sigma^2_{\mathcal{Y}^{(n)}} - 1). \quad (6)$$

### B. Closed-form expression of marginal laws

The expectations and variances of $X^{(n)}$ and $Y^{(n)}$ are noted respectively as $m_{X^{(n)}}$ and $m_{Y^{(n)}}$, $\sigma^2_{X^{(n)}}$ and $\sigma^2_{Y^{(n)}}$. We have:

$$m_{Y^{(n+1)}} - \mu_Y = \sigma_Y m_{\mathcal{Y}^{(n+1)}} = \rho^2(m_{Y^{(n)}} - \mu_Y). \quad (7)$$

The following expectation can then be obtained:

$$\begin{aligned} m_{X^{(n)}} &= \mu_X + \rho^{2n-1}\sigma_X(m_{Y^{(0)}} - \mu_Y)/\sigma_Y & (8)\\ m_{Y^{(n)}} &= \mu_Y + \rho^{2n}(m_{Y^{(0)}} - \mu_Y). & (9)\end{aligned}$$

Otherwise, $\sigma^2_{\mathcal{Y}^{(n+1)}} = \rho^4\sigma^2_{\mathcal{Y}^{(n)}} + 1 - \rho^4$ directly leads to:

$$\sigma^2_{Y^{(n+1)}} - \sigma^2_Y = \rho^4(\sigma^2_{Y^{(n)}} - \sigma^2_Y). \quad (10)$$

This results in the following variance:

$$\begin{aligned} \sigma^2_{X^{(n)}} &= \sigma^2_X + \rho^{4n-2}\sigma^2_X(\sigma^2_{Y^{(0)}} - \sigma^2_Y)/\sigma^2_Y & (11)\\ \sigma^2_{Y^{(n)}} &= \sigma^2_Y + \rho^{4n}(\sigma^2_{Y^{(0)}} - \sigma^2_Y). & (12)\end{aligned}$$

Similarly, the following covariance can be obtained:

$$\begin{aligned} \mathrm{Cov}(X^{(n)}, Y^{(n)}) &\triangleq E[(X^{(n)} - m_{X^{(n)}})(Y^{(n)} - m_{Y^{(n)}})] \\ = \rho\sigma_X\sigma_Y &\cdot \left[1 + \rho^{4n+2}(\sigma^2_{Y^{(0)}}/\sigma^2_Y - 1)\right]. & (13)\end{aligned}$$

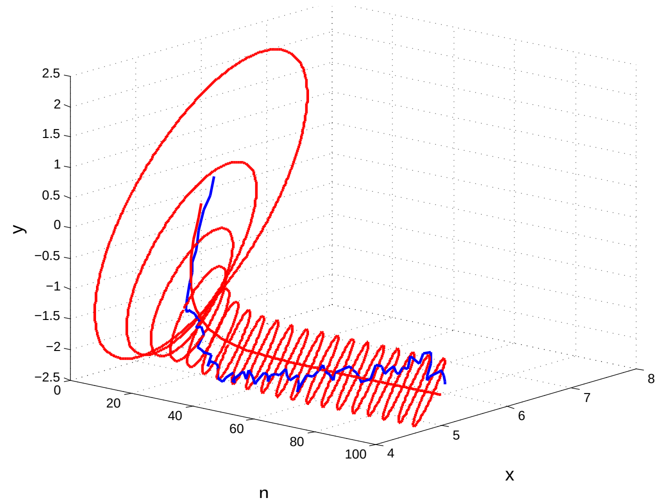


Fig. 1: Marginal distribution convergence to the target distribution ($\rho = 0.95$), as a function of the iteration $n$. In red, the mean and covariance ellipse every 10 iterations (with 95% confidence). In blue, a realisation of the Gibbs sampler.

Consequently, since all the distributions are normal, Eq. (8) and (11) state the convergence of the marginal normal distributions $f_{X^{(n)}}$ and $f_{Y^{(n)}}$, i.e. their first two moments, respectively towards the invariant target distribution $f_X$ and $f_Y$, when $n \to \infty$. It clearly depends on the correlation parameter $\rho$. The lower $|\rho|$ is, the more decorrelated the components $X$ and $Y$ are and the faster the convergence is. Conversely, the more correlated they are, the slower it converges. It also depends on the initial distribution. Tautologically, if one initiates with the target distribution, the convergence is already reached. Fig. 1 provides an illustration of the convergence of the marginal distributions $f_{X^{(n)},Y^{(n)}}$ to the target invariant normal distribution, closely linked to the "burn-in" period.

## III. Properties of the expectation estimator

We are interested in approximating expectations, such as $\mu_X = E(X)$ or $\mu_Y = E(Y)$. For the latter, we consider the usual MCMC estimate of Eq. (1):

### A. Bias determination

The expectation of $\bar{M}_Y$ is given by:

$$\begin{aligned} E(\bar{M}_Y) &= \frac{1}{\Delta N}\sum_{n=N_0}^{N} m_{Y^{(n)}} \\ &= \mu_Y + \frac{\sigma_Y}{\Delta N}\sum_{n=N_0}^{N} m_{\mathcal{Y}^{(n)}} \\ &= \mu_Y + \frac{\sigma_Y}{\Delta N} m_{\mathcal{Y}^{(0)}} \sum_{n=N_0}^{N} \rho^{2n} \\ &= \mu_Y + \frac{m_{Y^{(0)}} - \mu_Y}{\Delta N}\rho^{2N_0}\frac{1 - \rho^{2\Delta N}}{1 - \rho^2}. & (14)\end{aligned}$$

Noting $\alpha(\Delta N, \rho) \triangleq (1 - \rho^{2\Delta N})/[(1 - \rho^2)\Delta N]$, the bias $B(\bar{M}_Y) \triangleq E(\bar{M}_Y) - \mu_Y$ is then given by:

$$B(\bar{M}_Y) = \alpha(\Delta N, \rho) \cdot \rho^{2N_0} \cdot (m_{Y^{(0)}} - \mu_Y). \quad (15)$$

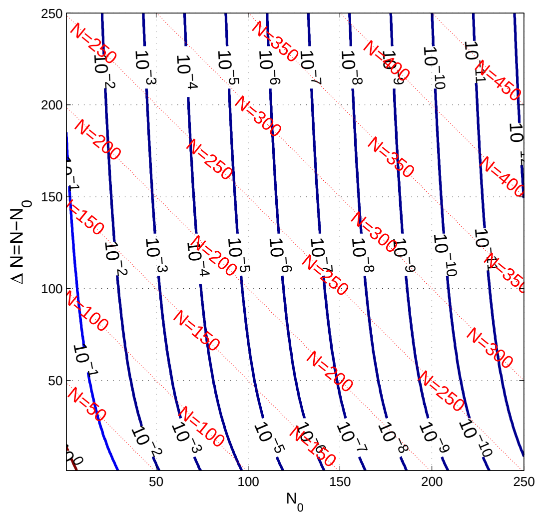


Fig. 2: Bias of $\bar{M}_Y$ ($\rho = 0.95$), function of $N_0$ and $\Delta N$. The straight isocontours of the sample budget ($N_0 + \Delta N$) are indicated by the red dotted line.

The bias also depends on the initial mean agreement $m_{Y^{(0)}} - \mu_Y$. Besides, the higher $|\rho|$ is and the more correlated the components are, the stronger the bias is. For a given $\rho$, it depends on $N_0$. A long "burn-in" phase reduces the bias. It also depends on $\Delta N$. One can notice that $\alpha(1, \rho) = 1$ and $\alpha(\Delta N, \rho) \stackrel{\Delta N \to \infty}{\longrightarrow} 0$. Furthermore, consistently with well-known general results [17, p. 21], the bias is $O(\Delta N^{-1})$ for high $\Delta N$ (for a given $N_0$); it will be further shown to be negligible for sufficiently long runs. The bias behaviour is illustrated in Fig. 2. For a given sample budget $N$ (in red), the optimal trade-off between $N_0$ and $\Delta N$ is straightforward. Concerning the bias, it is better to have a long "burn-in" period, discarding many samples and keeping only a few of the latest good ones.

### B. Variance determination

The variance of $\bar{M}_Y$ is given by:

$$\begin{aligned} V(\bar{M}_Y) &= E\left[\left(\bar{M}_Y - E(\bar{M}_Y)\right)^2\right] \\ &= \frac{1}{(\Delta N)^2} E\left[\left(\sum_{n=N_0}^{N-1} (Y^{(n)} - m_{Y^{(n)}})\right)^2\right] \\ &= \frac{1}{(\Delta N)^2} \sum_{n=N_0}^{N-1} \sum_{p=N_0-n}^{N-n-1} \gamma_Y^{n,p}, \end{aligned} \tag{16}$$

noting the lag-$p$ auto-covariance at time $n$

$$\gamma_Y^{n,p} \triangleq E\left[(Y^{(n)} - m_{Y^{(n)}})(Y^{(n+p)} - m_{Y^{(n+p)}})\right]. \tag{17}$$

It can be noticed that:

$$Y^{(n)} - m_{Y^{(n)}} = \sigma_Y \left(\mathcal{Y}^{(n)} - m_{\mathcal{Y}^{(n)}}\right). \tag{18}$$

Remark that $\gamma_Y^{n,p}$ can be rewritten as:

$$\begin{aligned} \gamma_Y^{n,p} &= \sigma_Y^2 E\left[\left(\mathcal{Y}^{(n)} - m_{\mathcal{Y}^{(n)}}\right)\left(\mathcal{Y}^{(n+p)} - m_{\mathcal{Y}^{(n+p)}}\right)\right] \\ &= \sigma_Y^2 \left(E\left[\mathcal{Y}^{(n)}\mathcal{Y}^{(n+p)}\right] - m_{\mathcal{Y}^{(n)}} m_{\mathcal{Y}^{(n+p)}}\right). \end{aligned} \tag{19}$$

For $p \geq 0$, it can be directly shown from Eq. (5) that:

$$E\left[\mathcal{Y}^{(n)}\mathcal{Y}^{(n+p)}\right] = \rho^2 E\left[\mathcal{Y}^{(n)}\mathcal{Y}^{(n+p-1)}\right], \tag{20}$$

ending as: $E\left[\mathcal{Y}^{(n)}\mathcal{Y}^{(n+p)}\right] = \rho^{2p} E\left[(\mathcal{Y}^{(n)})^2\right]$. Then, the auto-covariance recurrence can be written as:

$$\begin{aligned} \gamma_Y^{n,p} &= \sigma_Y^2 \rho^{2p} E\left[(\mathcal{Y}^{(n)})^2\right] - m_{\mathcal{Y}^{(n)}} m_{\mathcal{Y}^{(n+p)}} \\ &= \sigma_Y^2 \rho^2 \left(\rho^{2(p-1)} E\left[(\mathcal{Y}^{(n)})^2\right] - m_{\mathcal{Y}^{(n)}} m_{\mathcal{Y}^{(n+p-1)}}\right) \\ &= \rho^2 \gamma_Y^{n,p-1}. \end{aligned} \tag{21}$$

For any p, the auto-covariance is asymmetric:

$$\gamma_Y^{n,p} = \rho^{2|p|}\sigma_Y^2 + \rho^{4n+2p}(\sigma_{Y^{(0)}}^2 - \sigma_Y^2) \tag{22}$$

Notice that it tends to be symmetric when $n \to \infty$.

From Eq. (16), the $\bar{M}_Y$-variance can now be expressed as:

$$\begin{aligned} V(\bar{M}_Y) &= \frac{1}{(\Delta N)^2} \sum_{n=N_0}^{N-1} \left[\sigma_Y^2 \left(\sum_{p=N_0}^{N-n-1} \rho^{2p} + \sum_{q=1}^{n-N_0} \rho^{2q}\right)\right. \\ &\quad \left. + \rho^{4n}(\sigma_{Y^{(0)}}^2 - \sigma_Y^2) \sum_{p=N_0-n}^{N-n-1} \rho^{2p}\right]. \end{aligned} \tag{23}$$

A few geometric sums latter, it results in:

$$V(\bar{M}_Y) = \delta^2(\Delta N, \rho) \cdot \sigma_Y^2 + \alpha^2(\Delta N, \rho) \cdot \rho^{4N_0} \cdot (\sigma_{Y^{(0)}}^2 - \sigma_Y^2), \tag{24}$$

with notations:

$$\delta^2(\Delta N, \rho) \triangleq \alpha^2(\Delta N, \rho)\beta(\Delta N, \rho) \tag{25}$$

$$\beta(\Delta N, \rho) \triangleq \frac{\Delta N(1 - \rho^4) - 2\rho^2(1 - \rho^{2\Delta N})}{(1 - \rho^{2\Delta N})^2}. \tag{26}$$

The first factor is proportional to the variance $\sigma_Y^2$. The second one is very similar to the bias of Eq. (15), with a dependence on the initial variance agreement $\sigma_{Y^{(0)}}^2 - \sigma_Y^2$. For the special case of $\Delta N = 1$, notice that $\delta^2(1, \rho) = 1$ and $\alpha^2(1, \rho) = 1$. The variance $V(\bar{M}_Y)$ is then determined by $N_0$ and $\rho$, its lower bound being $\sigma_Y^2$.

Asymptotically, for high $\Delta N$ and a given $\rho$, it is straightforward to show that: $\delta^2(\Delta N, \rho) = O(\Delta N^{-1})$ and $\alpha^2(\Delta N, \rho) = O(\Delta N^{-2})$. That means that the second term and the bias of Eq. (15) are going to be negligible, compared to the first one. Let us stress that the convergence order $O(\Delta N^{-1})$ of the toy model variance $V(\bar{M}_Y)$ is consistent with Central Limit Theorems for Markov chains [1], [17]. On the other hand, it can be shown that $V(\bar{M}_Y)$ is non-zero when $|\rho| \to 1$.

The variance behaviour is illustrated in Fig. 3. The convergence order $O(\Delta N^{-1})$ of the variance is confirmed. For a given sample budget $N$ (in red), the optimal trade-off between $N_0$ and $\Delta N$ corresponds to the intersection of the variance level set and the budget line corresponding to $N$. It turns out to be always around $N_0 = 20$, whatever the value of $N$ is. Concerning the variance, it is better to have a short "burn-in" phase, discarding a few tens of samples and keeping all the next ones.

We obtain a closed-form expression of the variance $V(\bar{M}_Y)$ that provides a better comprehension. Let us stress that $V(\bar{M}_Y)$ cannot generally be calculated in closed-form, except for a few small problems; it must be estimated from simulations.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2020.3038136, IEEE Signal Processing Letters



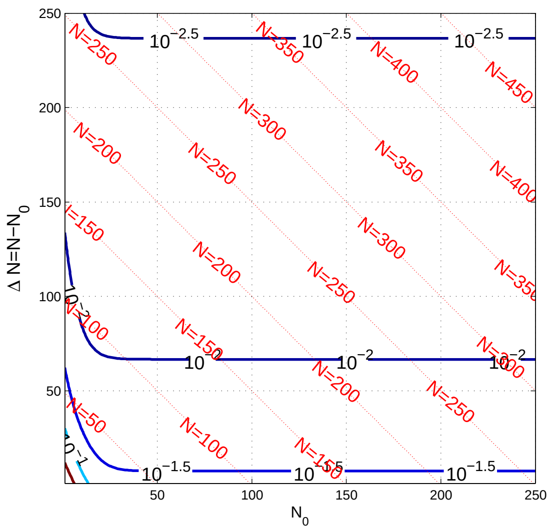


Fig. 3: Variance of $\bar{M}_Y$ ($\rho = 0.95$).

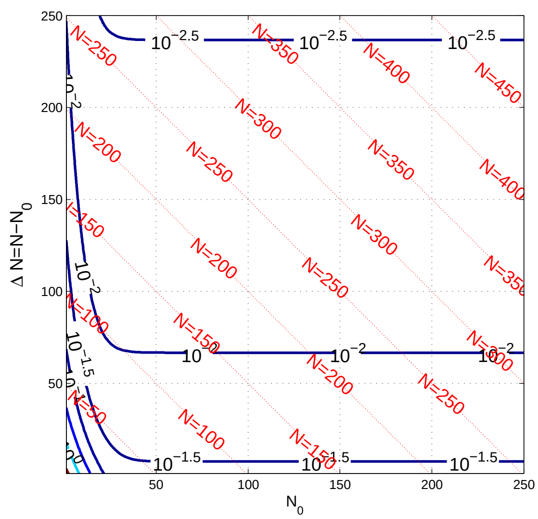


Fig. 4: Mean Squared Error of $\bar{M}_Y$ ($\rho = 0.95$).

### C. Mean Squared Error

The Mean Squared Error (MSE) of $\bar{M}_Y$ is given by:

$$\begin{aligned} \text{MSE}(\bar{M}_Y) &\triangleq V(\bar{M}_Y) + B(\bar{M}_Y)^2 \\ &= \delta^2(\Delta N, \rho) \cdot \sigma_Y^2 + \alpha^2(\Delta N, \rho) \cdot \rho^{4N_0} \cdot \\ &\quad \left[(\sigma^2_{Y(0)} - \sigma^2_Y) + (m_{Y(0)} - \mu_Y)^2\right]. \quad (27) \end{aligned}$$

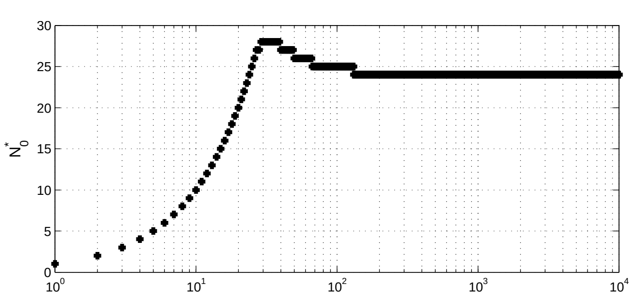


Fig. 5: Optimal "burn-in" length $N_0^\star$ ($\rho = 0.95$).

The second term depends on both the initial mean and variance agreements $(\sigma^2_{Y(0)} - \sigma^2_Y) + (m_{Y(0)} - \mu_Y)^2$. It must be noticed that: $\text{MSE}(\bar{M}_Y) = \sigma_Y^2 / \Delta N$ if $\rho = 0$. Asymptotically, for high $\Delta N$ and a given $\rho$, it is again straightforward to show that the second term in $O(\Delta N^{-2})$ is negligible, compared to the first one in $O(\Delta N^{-1})$.

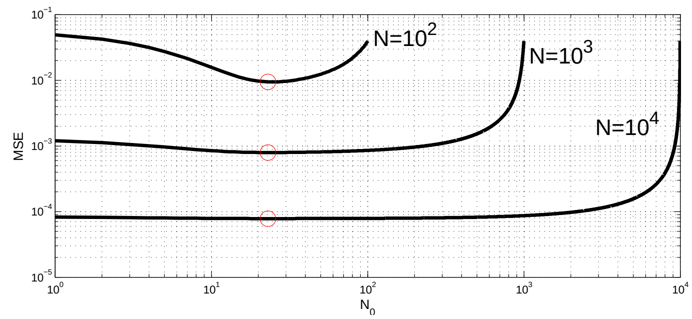


Fig. 6: MSE function of the "burn-in" length $N_0$ ($\rho = 0.95$). The optimal values are represented in red.

The MSE behaviour, illustrated in Fig. 4, is similar to the previous Fig. 3. For a given sample budget $N$ (in red), the trade-off between $N_0$ and $\Delta N$ seems to vary slightly, depending on $N$. The optimal value $N_0^\star$ is represented in Fig. 5, as a function of $N$. For low $N$, it is dominated by the bias reduction. It increases linearly up to $N \approx 30$, with the corresponding $\Delta N$ being constant and equal to 1. Thus, for low $N$, it is optimal, according to the MSE, to discard all the samples except the last one. For higher $N$, the optimal $N_0^\star$ slightly decreases and stabilises around 25 for high $N = 10^3$ to $10^6$. In addition, Fig. 6 represents the MSE as a function of $N_0$ for various values of $N$. In accordance with [17], the "burn-in" turns out to be useless, i.e. the curve is flat for high $N$.

## IV. Conclusion

This didactic investigation ends up with a reinforced understanding of Gibbs sampling and more generally, MCMC algorithms. Our non-asymptotical results are given in the context of a normal bivariate model. The Markov chain estimator is characterised, leading to a non-stationary analysis of the influence of the "burn-in" length. This theoretical work is new, even compared to recent works [11]–[15].

So, keep the samples or drop them? The trade-off is analysed in an original way. Roughly speaking, the answer is to have a long run, in which case the question is useless. Of course, there are limits to the transposition of toy models to real-world problems [17] where there are no closed-form expressions and where the target distribution is obviously not known. And yet, as in [4], it can provide useful guidelines to apply MCMC samplers.

## References

[1] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[2] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine learning*, vol. 50, no. 1-2, pp. 5–43, 2003.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 721–741, 1984.

[4] A. Gelman, G. O. Roberts, W. R. Gilks *et al.*, "Efficient Metropolis jumping rules," *Bayesian statistics*, vol. 5, no. 599-608, p. 42, 1996.

[5] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[6] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[7] S. N. MacEachern and L. M. Berliner, "Subsampling the Gibbs sampler," *The American Statistician*, vol. 48, no. 3, pp. 188–190, 1994.

[8] W. R. Gilks, G. O. Roberts, and E. I. George, "Adaptive direction sampling," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 43, no. 1, pp. 179–189, 1994.

[9] M.-H. Chen and B. Schmeiser, "Toward black-box sampling: A random-direction interior-point Markov chain approach," *Journal of Computational and Graphical Statistics*, vol. 7, no. 1, pp. 1–22, 1998.

[10] W. A. Link and M. J. Eaton, "On thinning of chains in MCMC," *Methods in ecology and evolution*, vol. 3, no. 1, pp. 112–115, 2012.

[11] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3123–3138, 2015.

[12] L. Martino, V. Elvira, and G. Camps-Valls, "The recycling Gibbs sampler for efficient learning," *Digital Signal Processing*, vol. 74, pp. 1–13, 2018.

[13] C. Fox, A. Parker *et al.*, "Accelerated Gibbs sampling of normal distributions using matrix splittings and polynomials," *Bernoulli*, vol. 23, no. 4B, pp. 3711–3743, 2017.

[14] A.-C. Barbos, F. Caron, J.-F. Giovannelli, and A. Doucet, "Clone MCMC: parallel high-dimensional gaussian Gibbs sampling," in *Advances in Neural Information Processing Systems*, 2017, pp. 5020–5028.

[15] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," *IEEE Transactions on Signal Processing*, vol. 67, no. 6, pp. 1648–1661, 2019.

[16] Y. Bar-Shalom and X.-R. Li, "Estimation and tracking- Principles, techniques, and software," *Norwood, MA: Artech House, Inc*, 1993.

[17] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.