Lab 12 of Thursday 4th December 2025

Exercice 1.

At every iteration of the general Metropolis–Hastings algorithm, a new candidate state Y_{n+1} is proposed by sampling $Y_{n+1} \sim q(X_n, \cdot)$, given the current state X_n. Here, $q(x, y)$ is the so-called proposal density. Consider now the case where the proposal does not depend on the current state, that is $q(x, y) \equiv q(y)$, so that the proposed candidate is $Y_{n+1} \sim q$. This particular Markov Chain Monte Carlo (MCMC) variant is sometimes called *independent Metropolis–Hastings algorithm* with fixed proposal (or simply *independence sampler*). Let's denote the target density by f. As such, this MCMC variant appears very similar to the Accept–Reject method for sampling from f (cf. Lab 02).

1) Suppose there exists a positive constant C such that $f(x) \leqslant Cq(x)$ for any $x \in \text{supp}(f) = \{x \in \mathbb{R}^d : f(x) > 0\}$. Show that the expected acceptance probability of the independent Metropolis–Hastings algorithm is *at least* $\frac{1}{C}$ whenever the chain is stationary. How does this compare to the expected acceptance probability of an Accept–Reject method?

2) Let us compare the independent Metropolis–Hastings algorithm and the Accept–Reject method in some more detail by an example. Specifically, the goal is to sample from a Gamma distribution with shape parameter α and scale parameter β, denoted by $\text{Gamma}(\alpha, \beta)$, so that the target PDF reads $f(x) \equiv f(x; \alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha) \mathbb{I}_{\{x \geqslant 0\}}$, where $\Gamma(\cdot)$ denotes the Gamma function.

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

   b) Use your Accept–Reject method to generate m random numbers $X_1, \dots, X_m$ with each $X_i \sim \text{Gamma}(\alpha, 1)$, when using $n = 5000$ random variables $Y_1, \dots, Y_n$ from the auxiliary $\text{Gamma}([\alpha], [\alpha]/\alpha)$ distribution. Notice that m is a random variable, which is smaller than n due to rejections. Perform the simulations for $\alpha = 4.85$.

   c) Implement the independent Metropolis–Hastings algorithm using as proposal q the PDF of the $\text{Gamma}([\alpha], [\alpha]/\alpha)$ distribution.

   d) Use the same sample $Y_1, \dots, Y_n$ used within the Accept–Reject method, now in the corresponding Metropolis–Hastings algorithm to generate $n = 5000$ realizations of the target distribution $\text{Gamma}(\alpha, 1)$ with $\alpha = 4.85$.

   e) Compare both methods with respect to:

      i. their acceptance rates,

      ii. their estimates for the mean of the $\text{Gamma}(4.85, 1)$ distribution, which is 4.85,

[1] *Hint: Recall that* $\sum_{k=1}^{K} \xi_k \sim \text{Gamma}(K, \beta)$ *for* $K \in \mathbb{N}$*, if* $\xi_k \overset{\text{i.i.d.}}{\sim} \text{Gamma}(1, \beta) \equiv \text{Exp}(\beta)$.

iii. the correctness of the target distribution,

Discuss your results.

Exercice 2.

Consider the Random Walk Metropolis–Hastings (RWMH) algorithm with proposal density $q(x,y) = g_\sigma(y-x)$ and target density $f\colon \mathbb{R} \to \mathbb{R}^+$. Let g_σ denote the density of the $\mathcal{N}(0,\sigma^2)$ distribution and suppose that

$$f(y) = \frac{1}{Z}\exp\left[-\left(\frac{1}{4}y^4 - \frac{1}{2}y^2 + \frac{1}{4}\right)\right] ,$$

where Z is such that f is a PDF on $\mathcal{X} = \mathbb{R}$. Suppose we wish to estimate $\mu = \mathbb{E}_f(\phi)$ for a suitable function $\phi\colon \mathbb{R} \to \mathbb{R}$. Let $\hat{\mu}_n^{\mathrm{MH}}$ be the estimator for μ based on the Markov chain of length n generated by the RWMH algorithm. Derive asymptotic confidence intervals for μ at probability level α using the CLT for Metropolis–Hastings Markov Chains. Within your simulations, estimate[2] this confidence interval and stop the Markov chain once the half-length of the interval is smaller than a given tolerance $\tau > 0$. Implement the following heuristics to estimate the required *time average variance constant* and compare their performance for different functions ϕ; namely $\phi(x) = x^p, p \in \mathbb{N}$. The time average variance constant is the asymptotic variance introduced in Theorem 8.10 of the lecture notes.

1) *Initial positive sequence estimator:*

$$\tilde{\sigma}^2 \approx \hat{\sigma}^2_{\mathrm{pos},n} := -\hat{c}_n(0) + 2\sum_{k=1}^{K}\left(\hat{c}_n(2k) + \hat{c}_n(2k+1)\right) ,$$

where K is the largest integer such that $\hat{c}_n(2k) + \hat{c}_n(2k+1) > 0$ for all $k = 1, \dots, K$. Here,

$$\hat{c}_n(j) := \frac{1}{n}\sum_{i=1}^{n-j}\left(\phi(X_i) - \hat{\mu}_n^{\mathrm{MH}}\right)\left(\phi(X_{i+j}) - \hat{\mu}_n^{\mathrm{MH}}\right)$$

is an appropriate covariance estimator in this context.

2) *Initial monotone sequence estimator:*

$$\tilde{\sigma}^2 \approx \hat{\sigma}^2_{\mathrm{mon},n} := -\hat{c}_n(0) + 2\sum_{k=1}^{K}\min_{1\leqslant j\leqslant k}\left\{\hat{c}_n(2j) + \hat{c}_n(2j+1)\right\} ,$$

where K and $\hat{c}_n$ are as for the initial positive sequence estimator above.

3) *Batch means estimator:* Suppose the Markov chain is $X_1, \dots, X_n$ at iteration n. Divide these n values into $N_b \in \mathbb{N}$ batches, each of length $N_\ell = n/N_b$. A typical decomposition is $N_\ell = n^{1-a}$ and $N_b = n^a$ for $a \in [0,1]$, for example $a = 0.5$, modulo integer rounding. Let

$$\hat{\mu}_i = \frac{1}{N_\ell}\sum_{j=(i-1)N_\ell+1}^{iN_\ell}\phi(X_j) , \quad i = 1, \dots, N_b ,$$

[2] The estimation has to be carried out on-the-fly, that is while the Markov chain evolves.

be the sample mean of the i-th batch. For N_ℓ sufficiently large, one can consider the batch means $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_{N_b}$ to be approximately mutually independent. Consequently, one can estimate the time average variance constant σ^2 by the sample variance estimator for independent realizations:

$$\tilde{\sigma}^2 \approx \hat{\sigma}^2_{\mathrm{BM},n} := \frac{n}{N_b} \frac{1}{N_b - 1} \sum_{i=1}^{N_b} \left(\hat{\mu}_i - \hat{\mu}_n^{\mathrm{MH}}\right)^2 .$$

In addition, instead of using all of the points in the Markov chain, one can as well discard a burn-in time B and replace $\sum_{k=1}$ with $\sum_{k=B}$ in the above estimators. Experiment with the effects of burn-in time on the above asymptotic variance estimators.

Exercice 3.

Consider a Markov chain $\{X_n\} \sim \text{Markov}(\pi, P)$ on a discrete state space $\mathcal{X}$ at equilibrium, with P irreducible, and π the unique invariant probability measure of P. Let l^2_π be the Hilbert space $l^2_\pi = \{f : \mathcal{X} \to \mathbb{R} : \sum_{i \in \mathcal{X}} f(i)^2 \pi_i < \infty\}$ with inner product $(f, g)_{l^2_\pi} = \sum_{i \in \mathcal{X}} f(i) g(i) \pi_i$, and $l^2_{\pi,0} = \{f \in l^2_\pi : \mathbb{E}_\pi[f] = 0\}$.

1) Show that if (P, π) are in detailed balance, then $(Pf, g)_{l^2_\pi} = (f, Pg)_{l^2_\pi}$ for any $f, g \in l^2_\pi$

2) Show that $\mathbb{E}[f(X_n) f(X_m)] = (P^{m-n} f, f)_{l^2_\pi}$ for any $f \in l^2_\pi$ and $m > n$.

3) Consider now the estimator

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^{N} f(X_n)$$

of $\mu = \mathbb{E}_\pi[f]$ under the assumption that $f \in l^2_\pi$. Show that $\mathbb{E}_\pi[\hat{\mu}_N] = \mu$, and

$$\mathbb{V}ar[\hat{\mu}_N] = \frac{1}{N} \sum_{l=0}^{N} c_l (P^l \tilde{f}, \tilde{f})_{l^2_\pi},$$

with $\tilde{f} = f - \mathbb{E}_\pi[f] \in l^2_{\pi,0}$ and

$$c_{l,N} = \begin{cases} 1, & l = 0 \\ 2(1 - \frac{l}{N}), & l > 0 \end{cases} \tag{3.1}$$

4) Conclude that the asymptotic variance $\mathbb{V}(f, p) := \lim_{N \to \infty} N \mathbb{V}ar_\pi(\hat{\mu}_N)$ satisfies $\mathbb{V}(f, p) = ((2(I - P)^{-1} - I)\tilde{f}, \tilde{f})_{l^2_\pi}$ if

$$\sup_{g \in l^2_{\pi,0}} \frac{(Pg, g)_{l^2_\pi}}{\|g\|_{l^2_\pi}} = \beta < 1. \tag{3.2}$$

5) Consider now the two irreducible transition matrices P_1 and P_2, both in detailed balance with π and satisfying (3.2) for some β_1, β_2. Show that if $(P_1)_{ij} \geqslant (P_2)_{ij} \forall i \neq j$, then

$$\mathbb{V}(f, P_1) \leqslant \mathbb{V}(f, P_2), \tag{3.3}$$

for any $f \in l^2_\pi$.

Hint: Take $P(\lambda) = (1 - \lambda) P_1 + \lambda P_2, \lambda \in [0, 1]$ and show that $\frac{d}{d\lambda} \mathbb{V}(f, P(\lambda)) \geqslant 0$.