

Class notes: Monte Carlo methods
Week 4, Markov chain Monte Carlo, auto-correlation
Jonathan Goodman
September 30, 2020

1 Auto-correlation, theory and practice

As before, suppose ρ is a target probability distribution of X and we are trying to estimate

$$B = E_\rho[V(X)] .$$

We have samples X_k from an MCMC sampler of ρ . The estimate is

$$\widehat{B}_N = \frac{1}{N} \sum_{k=1}^N V(X_k) . \quad (1)$$

The variance of the estimator will be written σ_N^2 :

$$\sigma_N^2 = \text{var}(\widehat{B}_N) . \quad (2)$$

Error bars for \widehat{B}_N come from estimates of σ_N^2 .

A practical question for this week is: how do we estimate σ_N^2 ? A theoretical question is: how do we prove that σ_N^2 goes to zero as $N \rightarrow \infty$? Our answers to these questions involves the steady state *auto-covariance* function

$$C_t = \text{cov}_\rho(V(X_0), V(X_t)) . \quad (3)$$

The notation cov_ρ means that we take the covariance under the assumption that $X_0 \sim \rho$. It is impossible in practice to take $X_0 \sim \rho$, but this mathematical quantity still is defined. It is meaningful because the *lag t* auto-covariance converges to C_t for any starting distribution (if the chain is non-degenerate)

$$\text{cov}_{\rho_0}(V(X_k), V(X_{k+t})) \rightarrow C_t \quad \text{as } k \rightarrow \infty . \quad (4)$$

This cov_{ρ_0} means that $X_0 \sim \rho_0$. In a practical code, ρ_0 , the starting distribution, may be that $X_0 = x_0$. This is a non-random start. Subsection ?? explains why we expect this to be true. Note, now, that this limit also applies to $-t$, because we can take $k' = k + t$ and then $X_k = X_{k'-t}$. The steady state auto-covariance function is a symmetric function of the lag, t :

$$C_{-t} = C_t .$$

The correlation between two random variables is a dimensionless version of the variance:

$$\text{corr}(U, W) = \frac{\text{cov}(U, W)}{\sqrt{\text{var}(U) \text{var}(W)}} . \quad (5)$$

The lag t *auto-correlation* is the steady state correlation between $V(X_k)$ and $V(X_{k+t})$. The general correlation formula simplifies because the distribution of $V(X_k)$ is the same as the distribution of $V(X_{k+t})$. This makes the two variance factors in the denominator of (5) equal to each other, and both are equal to $\text{var}_\rho(V(X_0)) = C_0$. The numerator is the auto-covariance. We denote the auto-correlation by D_t and have

$$D_t = \text{corr}_\rho(V(X_0), V(X_t)) = \frac{C_t}{C_0}. \quad (6)$$

We call the denominator the *static variance* because it depends only on ρ and not on the MCMC dynamics

$$C_0 = \text{var}_\rho(V(X)).$$

“Static” is the opposite of “dynamic”. Like the auto-covariance function, the auto-correlation function may be defined as a limit starting from any starting distribution and therefore is a symmetric function of t defined for all integers t :

$$D_t = \lim_{k \rightarrow \infty} \text{corr}_{\rho_0}(V(X_k), V(X_{k+t})).$$

The *auto-correlation time* (more properly, the *integrated* auto-correlation time) is

$$\tau = \sum_{t=-\infty}^{\infty} D_t = \sum_{t=-\infty}^{\infty} \frac{C_t}{C_0}. \quad (7)$$

This is the τ described in (12) of the Week 3 notes. For large N , we have

$$\sigma_N^2 \approx \frac{1}{N_{\text{eff}}} \text{var}_\rho(V(X)), \quad N_{\text{eff}} = \frac{N}{\tau}.$$

We will derive this formula in Section 5.

2 Partial resampling, Gibbs sampler

An MCMC *move* is a random function for constructing a Markov chain that preserves the target distribution ρ . A move can be expressed through its transition distribution R , or it can be expressed as a function with random inputs

$$X_{k+1} = S(X_k, \xi_k), \quad \xi_k \text{ i.i.d.} \quad (8)$$

The function $S(x, \xi)$ represents the actual MCMC code that takes the current state X_k and some independent random input ξ and produces the next state X_{k+1} . This preserves ρ if $X_k \sim \rho$ implies that $X_{k+1} \sim \rho$. We had the example of the *global move* Gaussian proposal Metropolis move last week. *Global* means that it moves all components of X .

It is possible to make an MCMC sampler by combining “local moves” that change just one or a few components of the random object $X = (X_1, \dots, X_n)$.

A *single variable move* would change only X_j , leaving the others the same. Suppose $\rho(x)$ is the target distribution. For each j , define x_j to be component j of x , and x_j^c to be the $n - 1$ component object with x_j left out. For example, if $x = (2, 3, 5, 7, 11)$, then $x_3 = 5$ and $x_3^c = (2, 3, 7, 11)$. The conditional density of x_j with the other components fixed is

$$\rho_j(x_j | x_j^c) = \frac{1}{Z(x_j^c)} \rho(x) .$$

The normalization factor $Z(x_j^c)$ is given by an integral that is usually impossible to compute in closed form. A *partial resampler* is a function that moves X_j in a way that preserves the conditional distribution of X_j given the remaining variables X_j^c . A partial resampler is a function $S_j(x_j, x_j^c, \xi)$ so that if $X_j \sim \rho_j(\cdot | x_j^c)$ and $\tilde{X}_j = S_j(X_j, X_j^c, \xi)$, then $\tilde{X}_j \sim \rho_j(\cdot | X_j^c)$. Partial resamplers may be practical because there are ways to sample one dimensional or low dimensional distributions. If you use a direct sampler to resample $X_j \sim \rho_j(\cdot, X_j^c)$, that is called [The *Gibbs sampler*, or the *heat bath*. The term “Gibbs sampler” is particularly unfortunate because it has nothing to do with any idea of the American physicist and mathematician J. Willard Gibbs, the first American physicist to achieve an “international” (European) reputation.] A partial resampler does not have to be a direct sampler. For example, it could be a one variable Metropolis move with a possibility of rejection.

It is “easy to see” that partial resampling preserves ρ . But a single partial resampling does not produce the target distribution. A *sweep* through the variables means

```
for j in range(n):           # sweep through all the components
    X[j] = resamp( j, X, rg) # resample X_j
```

One sweep changes every component. You can think of one sweep as being like one global metropolis move.

3 Discrete state space

The theory of MCMC is simpler to explain when the state space is finite and the MCMC “move” is given by an $n \times n$ stochastic matrix R . Most of the theory applies to infinite or continuous state space. The transition matrix has elements $R_{ij} = \Pr(i \rightarrow j) = \Pr(X_{k+1} = j | X_k = i)$. If $X_k \sim \rho_k$, and ρ_k is represented as a row vector, then $\rho_{k+1} = \rho_k R$. This formula may be iterated. If m is any positive integer, then

$$\rho_{k+m} = \rho_k R^m .$$

You find out what happens in m MCMC steps by taking the m -th power of the transition matrix. Eigenvalues and eigenvectors help to understand powers of matrices.

We call R a *stochastic matrix* if $R_{ij} \geq 0$ for all i, j , and if

$$\sum_{j=1}^n R_{ij} = 1, \quad \text{for all } i. \quad (9)$$

A transition matrix for a Markov chain is a stochastic matrix. The entries are non-negative because they are probabilities. The *row sums* (9) are equal to 1 because the sum is over all possible values of X_{k+1} .

3.1 Degenerate and non-degenerate Markov chains

Consider three transition matrices

$$R = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix} \quad (10)$$

This is an example of a discrete random walk on a *linear chain* with *reflecting boundary conditions* at the ends. The state X_k can go to $X_{k+1} = X_k \pm 1$ or $X_{k+1} = X_k$ with certain non-zero probabilities. If it tries to walk off an end from $X_k = 1$ to $X_{k+1} = 0$, or from $X_k = n$ to $X_{k+1} = n + 1$, that is rejected. Here $n = 5$.

$$R = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (11)$$

This consists of two blocks $\{1, 2\}$ and $\{3, 4\}$. This R preserves the uniform distribution $\rho = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. There is no path from $i = 1$ to $j = 3$.

$$R = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \quad (12)$$

This symmetric random walk on a ring of $n = 6$ states, this time with *periodic* boundary conditions. That means that a $1 \rightarrow 0$ transition becomes $1 \rightarrow n$ and $n \rightarrow n + 1$ becomes $n \rightarrow 1$. The “stay” transitions $i \rightarrow i$ are forbidden. The chain preserves the uniform distribution on $n = 6$ states.

The first example (10) is non-degenerate. The second one (11) is degenerate because it is not strongly connected. The third one (12) is degenerate because it is not acyclic.

A *path* (or *allowed path*) is a sequence of states $\mathcal{P} = (x_0, x_1, \dots, x_L)$ so that each transition $x_k \rightarrow x_{k+1}$ is “allowed” in the sense that the probability of this transition is not zero: $R_{x_k, x_{k+1}} > 0$. The length of \mathcal{P} is the number of transitions, which is L . In the chain (12), the path $\mathcal{P} = (3, 4, 3, 2, 1, 6)$ is a path of length 5 that connects $x_0 = 3$ to $x_5 = 6$. There is a path of length L from i to j if and only if $\Pr(X_L = j | X_0 = i) > 0$. This probability is

$$(R^L)_{ij} .$$

Therefore, there is a path of length L from i to j if and only if the corresponding matrix element is positive, $R_{ij}^L > 0$. A transition matrix (or a Markov chain) is non-degenerate if it is *strongly connected* and *acyclic*, which are defined here:

Strongly connected

We say that states i and j *communicate* if there is a path from i to j of some length and a path from j to i of some possibly different length. The state space $\mathcal{S} = \{1, 2, \dots, n\}$ may be divided into *connected components*, which are subsets of \mathcal{S} . Two states being in the same connected component means that i and j communicate. The chain is *strongly connected* if every pair of states communicates. That is, if i and j are any two states, then there is a path from i to j . The examples (10) and (12) are both strongly connected. The example (11) is not strongly connected. The connected components are $\{1, 2\}$ and $\{3, 4\}$.

Suppose that there is a target distribution ρ with $\rho_j \neq 0$ for all j . Note that if $\rho_j = 0$, we would be wise to leave j out of our state space because we want a Markov chain that never visits j . For example, in the chain (11), if $X_0 = 1$, which is the same as $\rho_0 = (1, 0, 0, 0)$, then $\rho_{j,k} = \Pr(X_k = j) = 0$ for $j = 3$ or $j = 4$ for all k . This means it is impossible that $\rho_k \rightarrow (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, which is the target distribution. You can check that $\rho R = R$, so ρ is invariant.

Acyclic

A path is a *loop* if it starts and ends at the same state, which is $x_L = x_0$. If \mathcal{P}_∞ and \mathcal{P}_∞ are two loops starting and ending at state i , then you can “compose” the loops to get a loop of length $L_1 + L_2$ that starts and ends at i . You do one loop and then the other. The chain is cyclic with cycle period $m > 1$ if every loop on state i has a length that is a multiple of m . If the chain is strongly connected, then every state has the same cycle period. [I’m not giving a proof, but you can see it’s true in examples.] The ring example (12) has cycle period 2. Every loop has an even number of hops.

If R has a cycle period $m > 1$ then it is unlikely that $\rho_k \rightarrow \rho$ as $k \rightarrow \infty$, even if ρ is an invariant distribution. If you start with $X_0 = i$, then $\rho_{i,k} = 0$ if k is not a multiple of m . This is because there are no paths from i to i of length k in that case. A ring with an even number of states has cycle period 2. A ring with an odd number of states is acyclic (check this).

Perron Frobenius and ergodic theorem

The Perron Frobenius theorem assumes that R is a stochastic matrix that is strongly connected and acyclic. There are two conclusions

- There is a unique distribution ρ that is invariant under R ($\rho R = \rho$).
- For any starting distribution ρ_0 , if $\rho_k = \rho_0 R^k$, then $\rho_k \rightarrow \rho$ as $k \rightarrow \infty$.

The examples (11) and (12) show that the second conclusion may not be true if R is not strongly connected or if R is cyclic.

The *ergodic theorem* has the same hypotheses as the Perron Frobenius theorem. The ergodic theorem states that if $V(x)$ is any “observable”, then

$$\widehat{B}_N \rightarrow B, \quad \text{as } N \rightarrow \infty. \quad (13)$$

4 Review of linear algebra

This section describes parts of linear algebra that we need this week. These relate to eigenvalues and eigenvectors of stochastic matrices and their relation to decay of covariance functions. Eigenvalues and eigenvectors are useful for describing powers of a matrix, the transition matrix in this case. Most of this material is familiar to most students, but few students have seen all of it and very few have seen it with the notation here. Linear algebra is an odd part of mathematics. On one hand, most of it seems simple. But it is powerful. In the decades since my first linear algebra, I think I have looked at something in a new way or used it in a different way every year.

Suppose A is an $n \times n$ matrix. A left eigenvector is a row vector l , with possibly complex entries, so that $lA = \lambda l$ for some possibly complex number λ . A right eigenvector is a column vector r with possibly complex entries so that $Ar = \lambda r$. A complex number λ is an eigenvalue of A if there is a non-zero left or right eigenvector corresponding to λ . The characteristic polynomial of A is $p(\lambda) = \det(A - \lambda I)$. If $p(\lambda) = 0$, then $A - \lambda I$ is singular, so there is a non-zero r with $(A - \lambda I)r = 0$. There also is a non-zero l so that $l(A - \lambda I) = 0$. In linear algebra, they say that being singular from the left is equivalent to being singular from the right, or that row rank is column rank. Therefore, λ is a “left eigenvalue” (there is a left eigenvector) if and only if λ is a “right eigenvector”. We just call these numbers “eigenvalues”. Every polynomial has at least one root, so every matrix has at least one eigenvalue.

The matrix A is *diagonalizable* if there are n linearly independent right eigenvectors (or, equivalently, left eigenvectors). You might have the impression that “most matrices” are diagonalizable, but there are matrices that are not diagonalizable. For example,

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

You can check that if $Ar = \lambda r$, then

$$\lambda = 1, \text{ and } r = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \text{ for some } \alpha.$$

Thus, you cannot find two linearly independent right eigenvectors of A . We will see that the transition matrix of a Markov chain is diagonalizable if the chain satisfies detailed balance, but possibly not otherwise.

You can understand upper Schur form by multiplying both sides by Q and then looking at what the equation means in terms of vectors.

4.1 Upper Schur form

If a matrix is diagonalizable, then there is a basis in which the matrix is diagonal. If a matrix is not diagonalizable then this is impossible, by definition. Even if a matrix is diagonalizable, the eigenvectors may be “almost linearly dependent” in a sense we may come back to later in the course. If you base a mathematical discussion on eigenvectors, then you constantly have to say “if it’s diagonalizable, then \dots , otherwise \dots ”. The *upper Schur form* of a matrix is a substitute for a diagonal form that is useful because every matrix has one and because the *Schur vectors* are ortho-normal (definition below).

Suppose $q_k \in \mathbb{C}^n$ is a collection of n column vectors with possibly complex entries. These vectors form the columns of an $n \times n$ matrix

$$Q = \begin{pmatrix} \vdots & \vdots & \cdots & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix}.$$

If A is any matrix, then the *adjoint* matrix, or the *conjugate transpose* is the matrix formed by taking the transpose and the complex conjugate of each entry. We denote it by A^* .

$$A_{ij}^* = \overline{A_{ji}}.$$

Suppose r is an $n \times 1$ matrix, which may be thought of as a column vector,

$$r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}.$$

Then the following matrix product is a 1×1 matrix, which may be thought of as a number

$$r^* r = (\overline{r_1} \quad \overline{r_2} \quad \cdots \quad \overline{r_n}) \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \overline{r_1} r_1 + \cdots + \overline{r_n} r_n = \sum_{j=1}^n |r_j|^2.$$

This quantity is the square of the *norm* (more completely, the l^2 norm) of r

$$\|r\|^2 = r^* r .$$

A vector r is *normalized* if $\|r\| = 1$. Vectors r and s are *orthogonal* if $r^* s = 0$. The family of vectors q_j are *orthonormal* if

$$q_i^* q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j . \end{cases}$$

The entries of $Q^* Q$ are

$$(Q^* Q)_{ij} = q_i^* q_j .$$

Therefore, the vectors q_j are orthonormal if and only if Q has the property that

$$Q^* Q = I .$$

This means that $Q^* = Q^{-1}$. Any matrix commutes with its inverse, so $Q^* Q = I$ is equivalent to $Q Q^* = I$. A complex matrix with this property is called *unitary*. A real matrix with this property is called *orthogonal*.

A unitary matrix Q puts A into *upper Schur form* if there is an upper triangular matrix U so that

$$A = Q U Q^* .$$

The diagonal entries of U turn out to be the eigenvalues of A . The upper Schur form has a lot in common with the Jordan canonical form, except that the Jordan eigenvectors and generalized eigenvectors do not have to be ortho-normal, and different matrices have Jordan canonical forms with different Jordan structures. [Ignore these statements if you don't know Jordan form.]

$$\begin{aligned}
 A Q &= Q U \\
 A \begin{pmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & & \vdots \end{pmatrix} &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & q_n \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} \lambda_1 & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & \lambda_2 & u_{23} & & u_{2n} \\ 0 & 0 & \lambda_3 & & \vdots \\ \vdots & & & \ddots & \\ 0 & \cdots & & 0 & \lambda_n \end{pmatrix} \\
 \begin{pmatrix} \vdots & \vdots & & \vdots \\ A q_1 & A q_2 & \cdots & A q_n \\ \vdots & \vdots & & \vdots \end{pmatrix} &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ \lambda_1 q_1 & \lambda_2 q_2 + u_{12} q_1 & \cdots & \lambda_n q_n + u_{1n} q_1 + \cdots + u_{n-1,n} q_{n-1} \\ \vdots & \vdots & & \vdots \end{pmatrix}
 \end{aligned}$$

This says that

$$A q_j = \lambda_j q_j + (\text{linear combination of } q_i \text{ for } i < j) .$$

This is

$$A q_1 = \lambda_1 q_1$$

and

$$Aq_2 = \lambda_2 q_2 + u_{12} q_1$$

and

$$Aq_3 = \lambda_3 q_3 + u_{13} q_1 + u_{23} q_2$$

and so on.

4.2 Detailed balance and Rayleigh quotients

Practitioners like detailed balance because it gives correct MCMC methods. Theoreticians like detailed balance because the transition matrix is self adjoint. This leads to real eigenvalues and orthogonal eigenvectors, in the weighted inner product.

5 Auto-correlation time

6 Exercises

1. Suppose R_1 and R_2 are two “moves” ($n \times n$ stochastic matrices that preserve ρ). Give an example to show that a “sweep” that does these two moves might not have detailed balance even though R_1 and R_2 do have detailed balance. *Hint.* The product of symmetric matrices does not have to be symmetric.
2. Suppose R is an $n \times n$ stochastic matrix that is non-degenerate with respect to ρ . Suppose $V(x)$ is an observable and C_t is the lag t auto-covariance function. Show that $C_t \geq 0$ if t is even. Show that

$$\tau \leq \frac{2}{g} \quad \text{where } g \text{ is the spectral gap.}$$

3. You can calculate everything explicitly for the linear auto-regressive process

$$X_{k+1} = aX_k + bZ_k, \quad \text{where } Z_k \sim \mathcal{N}(0, 1), \quad \text{i.i.d.}$$

Assume $|a| < 1$. We saw last week that the invariant distribution is $\rho = \mathcal{N}(0, v)$, with a specific formula for v in terms of a and b . This exercise is practice with geometric sums. It explains the intuition that the auto-correlation time is a characteristic decay time for the MCMC process.

- (a) Take $V(x) = x$ and calculate the auto-covariance C_t . *Hint.* You can write $X_t = a^t X_0 + \text{independent random variables}$.
- (b) Calculate the sums of geometric series necessary to compute τ .
- (c) Show that $\tau > 0$ and decide whether $\tau < 1$ is possible.

(d) Now assume $X_0 = 0$ and find a formula for w_{kN} so that

$$\widehat{B}_N = \frac{1}{N} \sum_{k=0}^{N-1} w_{kN} Z_k .$$

This involves more geometric series.

(e) Explain and calculate (approximately if necessary)

$$\sigma_N^2 = \sum_{k=0}^{N-1} w_{kN}^2 .$$

Show that this is approximately the same as $v \frac{\tau}{N}$, where τ is from part (b).

(f) We saw that the auto-covariance function is a geometric series, so C_t is an exponential function of t for $t > 0$. The *exponential decay time*, or the *exponential auto-correlation time* is that τ_{exp} so that

$$a^{\tau_{\text{exp}}} = e^{-1} .$$

This is the “time” (the number of iterations) needed for a^t to decay from $a^0 = 1$ to $a^{\tau_{\text{exp}}} = \frac{1}{e}$. For this part only, assume $0 < a < 1$. Show that the exponential and the integrated auto-correlation times both go to infinity as a goes to 1 and show that one is approximately proportional to the other in that limit.

(g) (*extra credit*) Find the auto-correlation times for the observables $V(x) = x^n$ and show that these are related to the eigenvalues of the linear auto-regressive process from Exercise 1 of Week 3.

4. Suppose $\rho_0(x)$ is a probability density and we want to sample ρ_0 with a constraint that $X \in A$, for some set of allowed values A . That means that the target density is

$$\rho(x) = \frac{1}{Z} \begin{cases} \rho_0(x) & \text{if } x \in A \\ 0 & \text{if } x \notin A . \end{cases}$$

Suppose there is a direct sampler that produces i.i.d. samples from the unconstrained density ρ_0 . The rejection algorithm is explained in this not quite complete piece of code.

```
while True:
    X = rho_0_Samp(rg)    # generate a proposal
    if ( X is in A ):    # need actual code for this test
        break
```

Show that this is a direct sampler of ρ . Show that the expected number of proposals (calls to `rho_0_Samp`) is equal to $\Pr_{\rho_0}(X \in A)$. This is an example of *rejection sampling* from the *proposal distribution* ρ_0 .

5. A Gaussian *random walk* is a *path process* that describes a random path X with n steps $X = (X_1, \dots, X_n)$ that has independent Gaussian steps. We *normalize* by assuming that the steps have mean zero and variance 1. Step j is $X_{j+1} - X_j$. If X_0 , then the PDF of X is a multi-variate normal. The random walk is *pinned* if $X_{n+1} = 0$. The random walk is *one sided* if $X_j \geq 0$ for all j .

- (a) Write a formula for the PDF of a Gaussian random walk with n steps. Call this ρ_0 .
- (b) Write a formula for the PDF of a pinned random walk with n steps. One way to do this is to write a formula for a random walk with $n+1$ steps and then take the conditional density with the condition that $X_{n+1} = 0$.
- (c) Let \tilde{X} be a Gaussian random walk with $n+1$ steps and define

$$X_j = \tilde{X}_j - \frac{j}{n+1} \tilde{X}_n .$$

Show that X is a pinned Gaussian walk with n steps. Use this to describe a direct sampler for pinned Gaussian walks.

- (d) Write a formula for the PDF of a pinned and one sided walk with n steps ($X_{n+1} = 0$). Note that the normalization constant is unknown. It is related to the probability that a pinned walk is one-sided, which there is (I believe) no formula for. Call this distribution $\rho(x) = \rho(x_1, \dots, x_n)$.
6. This computational exercise involves two samplers for one-sided pinned random walks. An interesting statistic is the mean “excursion”

$$V(x) = \frac{1}{n} \sum_{j=1}^n x_j .$$

A phenomenon called *entropic repulsion* makes this quantity larger than one might at first guess. Use two MCMC methods. One is a global Gaussian proposal Metropolis strategy. The other is a sweep from $j = 1$ to $j = n$, where at each step you resample the single variable X_j with the correct conditional distribution. Do the resampling using rejection sampling (Exercise 4) from the appropriate Gaussian proposal distribution. Make a histogram of the values of $V(X)$ for each MCMC method to see that they agree to sampling error. You can do this with a small value of n (but not too small) because it’s a test for correctness of the two samplers. Once you believe the samplers are correct (or at least consistent with each other), do some experiments with larger n . Compute and plot the estimated auto-covariance function for the two methods and for various values of n . Experiment with the proposal size for the global Metropolis algorithm to see what size works well and how that size depends on n . Comment on

the decay of the auto-covariance functions for various n values. Comment on which method seems better. Make sure you do long enough MCMC runs for the estimated auto-covariance functions to be somewhat accurate.