Assignment 2.

Corrections: none yet.

1. (*The mathematics of detailed balance*) The more abstract view of transition matrices and detailed balance is helpful in some more complicated applications. If you are not used to abstract linear algebra, inner products and so forth, the Linear Algebra book by Peter Lax may help. Depending on your background, some of the steps will be very easy, or really confusing.

   (a) Suppose H is a real vector space, which you can usually think of as $\mathbb{R}^n$. An *inner product* on H is a function that produces a real number from any two vectors $u \in H$ and $v \in H$. We write this as $(u,v) \rightsquigarrow \langle u,v\rangle$. [The $\rightsquigarrow$ symbol means that we start with the pair, (u,v), and get the number, which is written $\langle u,v\rangle$. For example a mapping that gives the exponential probability density for parameter λ and t might be written $(\lambda,t) \rightsquigarrow \lambda e^{-\lambda t}$.] A real inner product is defined by the following axioms

   i. $\langle u,v\rangle = \langle v,u\rangle$ (*symmetry*).
   ii. If a_1 and a_2 are real numbers and $u_1 \in H$ and $u_2 \in H$, then $\langle (a_1u_1 + a_2u_2), y\rangle = a_1\langle u_1,v\rangle + a_2\langle u_2,v\rangle$. (*bi-linearity*, by symmetry it is linear in the v variable too.)
   iii. $\langle u,u\rangle \geq 0$ for all $u \in H$, and $\langle u,u\rangle = 0$ only if u 0 in H. (*positive definiteness*)

   An example is $H = \mathbb{R}^n$ and $\langle u,v\rangle = u^tv$. Now suppose $\mathcal{S}$ is a finite state space with n elements, written $x \in \mathcal{S}$, and $f(x)$ is a probability distribution on $\mathcal{S}$. Let H be the set of all real valued functions of $x \in \mathcal{S}$. Define

   $$\langle u,v\rangle = \mathrm{E}_f[u(X),v(X)] = \sum_{x\in\mathcal{S}} u(x)v(x)f(x)\ . \tag{1}$$

   Show that this is an inner product in that it satisfies the three properties above.

   (b) Let A be a linear transformation on H. The adjoint of A with respect to the inner product $\langle\cdot,\cdot\rangle$ will be written A^*. If A is given by a matrix, we write A^t for the matrix transpose. The adjoint is defined by the formula

   $$\langle u, Av\rangle = \langle A^*u, v\rangle\ ,$$

which is supposed to hold for every $u \in H$ and $v \in H$. It is easy (though too time consuming for this assignment) to show that for any A and any inner product there an A^* that satisfies this formula, and A^* is unique. The adjoint depends on A and on the inner product. If $H = \mathbb{R}^n$ and $\langle u, v\rangle \; u^t v$, then $A^* = A^t$. Suppose $v = Au$ is given by

$$v(x) = \sum_{x \in \mathcal{S}} a_{xy} u(y) \ .$$

and the inner product is given by (1). Let A^* be given by

$$v = A^* u \iff v(x) = \sum_{x \in \mathcal{S}} a^*_{xy} u(y) \ .$$

Find a formula for a^*_{xy} in terms of the entries of A and the probabilities f.

(c) Suppose the numbers P_{xy} represent the transition probabilities for a Markov chain on state space $\mathcal{S}$. That is,

$$P_{xy} = \mathrm{P}(x \to y) = \mathrm{P}(X_{n+1} = y \mid X_n = x) \ .$$

The *generator* of the Markov chain is the linear transformation defined on H defined by $v = Lu$ if

$$v(x) = \mathrm{E}[\, u(X_{n+1}) \mid X_n = x] \ . \tag{2}$$

Show that the matrix that represents the linear transformation L has entries given by the transition probabilities P_{xy}. [The definition (2) may seem ridiculously and pointlessly abstract, but it is the simplest way to express some more complicated Markov chains, particularly if the state space is continuous rather than discrete.]

(d) Let P be any transition matrix, f any probability distribution on $\mathcal{S}$, and u any function on $\mathcal{S}$. Suppose $X_n \sim f$. Let $\mathbf{1} \in H$ be the function that is always equal to 1, which means $\mathbf{1}(x) = 1$ for all $x \in \mathcal{S}$. Show that

$$\mathrm{E}_f[\, u(X_n)] = \langle u, \mathbf{1}\rangle \ .$$

Show that if $v = Pu$ as defined in (2), then

$$\mathrm{E}_f[\, u(X_{n+1})] = \langle v, \mathbf{1}\rangle \ .$$

Show that the definition (2) implies that $P\mathbf{1} = \mathbf{1}$. Use this to show that if $P = P^*$, and $X_n \sim f$, then, for any $u \in H$,

$$\mathrm{E}[\, u(X_{n+1})] = \mathrm{E}[\, u(X_n)] \ .$$

Conclude that $P = P^*$ implies that P preserves f.

(e) Suppose $\mathcal{S} = \mathbb{R}^2$. In an attempt to avoid confusion with indices let denote $x \in \mathbb{R}^2$ in components as $x = (y_1, y_2)$. Let $f(x) = f(y_1, y_2)$ be a probability density on $\mathcal{S}$. Consider partial resampling of Y_1 with Y_2 fixed. Suppose $u(x) = u(y_1, y_2)$ is a function on $\mathcal{S}$. Find an integral formula for $v(y_1, y_2)$, with $v = Pu$ defined by (2). Define the inner product as $\langle u, v\rangle = \int u(x)v(x)f(x)\,dx$. Show that $P^* = P$. Note that a direct formula for P_{xy} as an integral kernal

$$v(x) = \int_{\mathbb{R}^2} P_{xx'} u(y_1', y_2') dy_1' dy_2'$$

would involve delta functions because y_2 doesn't change.

2. (*The eigenvalue problem for self adjoint linear transformations*) Suppose H is a real inner product space as above and A is a linear transformation with $A^* = A$. Such a transformation is called *self adjoint* with respect to the inner product $\langle \cdot, \cdot \rangle$. The eigenvalue problem for a self adjoint linear transformation in a real inner product space is more or less the same as it is for symmetric matrices in $\mathbb{R}^n$.

(a) (*Eigenvalues are real, there are real eigenvectors.*) Suppose $A(u + iv) = (\lambda + i\mu)(u + iv)$. Here λ and μ are real numbers, and $u \in H$ and $v \in H$ are real vectors. Show that if $(u + iv) \neq (0 + i0)$, then $\mu = 0$. Show that u and v are real eigenvectors, at least one of which is not zero. Hint: calculate $\langle (u - iv), A(u + iv)\rangle$.

(b) (From here on, all quantities are real.) Show that if $Au = \lambda u$ and $Av = \mu v$ and $u \neq 0$ and $v \neq 0$, and $\lambda \neq \mu$, then $\langle u, v\rangle = 0$.

(c) The *Rayleigh quotient* is the function

$$v \rightsquigarrow R(v) = \frac{\langle v, Av\rangle}{\langle v, v\rangle} .$$

This is defined for any $u \neq 0$. Assume that u maximizes the Rayleigh quotient: $R(u) = \lambda = \max_{v\neq 0} R(v)$. Show that u is an eigenvector of A with eigenvalue λ. Hint: Show that for any $v \in H$, $\frac{d}{dt}R(u + tv) = \langle v, w\rangle$ when $t = 0$, with $w \neq 0$ if $Au \neq \lambda u$.

(d) Suppose u maximizes R as above, and $\langle v, u\rangle = 0$. Show that $\langle v, Au\rangle = 0$. Hint: if $\langle v, Au\rangle \neq 0$, then u was not a maximizer.

(e) Let $H^\perp \subset H$ be the set of vectors perpendicular to u: $v \in H^\perp$ if $\langle v, u\rangle = 0$. Show that the maximizer of R over $H^\perp$ is also an eigenvector of A. Conclude by induction that if H is n dimensional, then there is a basis of H consisting of n orthogonal eigenvectors of A.

(f) Suppose the eignevectors and eigenvalues are $Au_j = \lambda_j u_j$. Show that any $v \in H$ may be written as

$$v = \sum_{j=1}^{n} w_j u_j \;\; , \quad w_j = \frac{\langle u_j v\rangle}{\langle u_j, u_j\rangle} .$$

Show that

$$A^t v = \sum_{j=1}^{n} \lambda_j^t w_j u_j \ .$$

(g) (*The point of this exercise*) Suppose that P satisfies detailed balance with respect to probability distribution f. The *spectral gap* is

$$g = \min_{\lambda \neq 1} |1 - \lambda| \ .$$

The Perron Frobenius theorem states that if P is acyclic and irreducible, then $g > 0$. Let $\rho(t)$ be the lag t auto-correlation for an observable $v(X)$. Show that $|\rho(t)| \leq (1-g)^t$. Conclude that the auto-correlation time satisfies

$$\tau \leq \frac{2}{g} \ .$$

This says that $\frac{1}{g}$ is the worst case auto-correlation time scale.

(h) (*Another interesting consequence*) Show that if P satisfies detailed balance then $\rho(t) \geq 0$ whenever t is even. If the computed auto-corelation function has $\widehat{\rho}(t) < 0$, then the estimation error is at least $|\widehat{\rho}(t)|$.

3. (*Optimal acceptance probability.*) Suppose the state space is the non-negative integers $\mathbb{Z}_+ = \{0, 1, \ldots\}$. Suppose the target probability distribution is $f(x) = \frac{1}{Z} a^x$, for $x \in \mathbb{Z}_+$, and $0 < a < 1$. The normalization constant is easy to compute: $Z = \frac{1}{1-a}$, but this should be irrelevant to this problem. Consider the MCMC algorithm whose proposal if to move a "random" amount up to L steps in either direction. More precisely, the proposed step is $X \to Y$, where $Y \neq X$, and $|X - Y| \leq L$, and all of the $2L$ allowed Y values equally likely. The MCMC algorithm follows this proposal with a Metropolis test. Proposals with $Y < 0$ are always rejected. For any a, there is an optimal L. If a is close to 1 (.95, say) then $L = 1$ proposals are very likely to be accepted, but $L = 2$ proposals move X more and therefore have smaller auto-correlation time. If L is too large, then it is likely to propose Y values that are much larger than X (therefore likely to be rejected), or $Y < 0$, (definitely rejected).

For computer analysis, you should restrict the state space to $x < N$. But it will be possible to choose N so large that it does not effect the results, unless a is very close to 1. Restricting the state space changes the normalization constant, but that should not matter. In Python, compute the transition matrix, P, for the MCMC algorithm and use numpy to compute its eigenvalues and spectral gap. Also compute the acceptance probability. For a range of a values, look for the acceptance probability of the L value that maximized the spectral gap. Make one or two interesting plots that give some insight into this problem. Is there an optimal acceptance probability for this problem?

Programming requirements. The code must be general and automated, so that you can change the value of a single variable, run the program, and do all the computations and make annotated plots. The code should be commented and clearly written. Please hand in the code with the plots and your description of the conclusions.