

Optimization Conscious Econometrics Pset 2

Part 1

Timothy Schwieg

1 Exercise 5

Show that the Gibbs sampler satisfies detailed balance

Detailed balance requires that

$$P(\theta^{i-1}|\theta^i)p(\theta^i) = P(\theta^i|\theta^{i-1})p(\theta^{i-1})$$

Consider a single step of the Gibbs-Sampler on dimension j for θ .

Note that $\theta_{-j}^i = \theta_{-j}^{i-1} := \theta_{-j}$ as we are only sampling on a single dimension. There is no chance of rejection, so we only have to consider the full j^{th} conditional of the distribution.

$$\begin{aligned} p(\theta_j^{i-1}|\theta_{-j}) \Pr(\theta^i|\theta^{i-1}) &= p(\theta_j^{i-1}|\theta_{-j})p(\theta_j^i|\theta_{-j}^i) \\ &= p(\theta_j^{i-1}|\theta_{-j}) \frac{p(\theta_j^i|\theta_{-j})}{\int p(z|\theta_{-j})} \\ &= p(\theta_j^i|\theta_{-j}) \frac{p(\theta_j^{i-1}|\theta_{-j})}{\int p(z|\theta_{-j})} \\ &= p(\theta_j^i|\theta_{-j}) \Pr(\theta^{i-1}|\theta^i) \end{aligned}$$

2 Exercise 7

Give the expression for the exact distribution, called the *ABC* posterior which is sampled by the ABC sampler.

Drawing from Forneron and Ng (2016):

First note that we draw our possible value θ^* from $q(\theta^*|\theta^i)$, and then independently sample $z \sim F_Z$. So our proposal distribution is the product of these two:

$$q(\theta^*, z|\theta^i) = q(\theta^*|\theta^i)L(z|\theta^i)$$

Our acceptance probability is given by:

$$\alpha(\theta^i, \theta^{i+1}) = \min\left\{\mathbb{1}_{\{d(\hat{Y}, Y_{\theta^*})\}} \frac{q(\theta^i|\theta^{i+1})}{q(\theta^{i+1}|\theta^i)}, 1\right\}$$

By construction, any accepted θ by the distribution must satisfy $d(\hat{Y}, Y_\theta) < \epsilon$ and therefore $\mathbb{1}_{\{d(\hat{Y}, Y_\theta) < \epsilon\}} = 1$ so dividing by it will not affect the acceptance probability.

Manipulating the interior of the acceptance probability:

$$\begin{aligned} \mathbb{1}_{\{d(\hat{Y}, Y_{\theta^*}) < \epsilon\}} \frac{q(\theta^i | \theta^*)}{q(\theta^* | \theta^i)} &= \frac{\mathbb{1}_{\{d(\hat{Y}, Y_{\theta^*}) < \epsilon\}} q(\theta^* | \theta^i) L(z^* | \theta^i) L(z^i | \theta^*)}{\mathbb{1}_{\{d(\hat{Y}, Y_{\theta^i}) < \epsilon\}} q(\theta^i | \theta^*) L(z^* | \theta^i) L(z^i | \theta^*)} \\ &= \frac{\mathbb{1}_{\{d(\hat{Y}, Y_{\theta^*}) < \epsilon\}} q(\theta^i, z^i | \theta^*) L(z^* | \theta^i)}{\mathbb{1}_{\{d(\hat{Y}, Y_{\theta^i}) < \epsilon\}} q(\theta^*, z^* | \theta^i) L(z^i | \theta^*)} \end{aligned}$$

This is now a special case of the Metropolis Hastings algorithm, with a particular p distribution, so the detailed balance equations must be satisfied. This particular p must be the ABC posterior that is being sampled, and is given by:

$$p_{ABC}(\theta^i, z^i | \epsilon) = L(z^i | \theta^i) \mathbb{1}_{\{d(\hat{Y}, Y_{\theta^i}) < \epsilon\}}$$

3 Exercise 8

Show that the set of left and right eigenvalues are equal.

The space of right eigenvalues are the set of numbers $\lambda \in \Lambda_R$ that satisfy:

$$Pv = \lambda v \Leftrightarrow (P - \lambda I)v = 0$$

for some vector $v \neq 0$. This is equivalent to the set of numbers Λ_R that give non-trivial basis vectors to the Null Space of $P - \lambda I$. These are the numbers where $|P - \lambda I| = 0$.

The left eigenvalues can be defined in a similar manner: $\lambda \in \Lambda_L$ if it satisfies

$$vP = \lambda v \Leftrightarrow v(P - \lambda I) = 0 \Leftrightarrow (P - \lambda I)'v' = 0$$

for some vector $v \neq 0$. Again this set can be characterized by a determinant condition: $|(P - \lambda I)'| = 0$. However, this is just the transpose of the matrix for the right-eigenvalues, and the determinant of the transpose of a matrix is equal to the determinant of the matrix itself. This means that the two problems are equivalent, and therefore the set of left and right eigenvalues are equal.

4 Exercise 9

Write the computation of Total Variance as an optimal transport problem. In particular, display explicitly the cost matrix.

We know that Total Variance is defined as:

$$\|\mu - \nu\| = \max_{A \subset X} |\mu(A) - \nu(A)| = \inf_P \{\Pr(X \neq Y) \mid (X, Y) \text{ is a coupling with joint distribution } P\}$$

This is the largest possible difference in the probabilities that the two measures can assign to the same event.

Following the questions suggested in the email:

1. For a given point $y \in Y$, how much mass in total must have been moved to y ?

We require that the $\nu(y)$ mass be moved into Y at the point y to ensure that this distribution has the “correct” mass.

2. For a given point $x \in X$ how much mass must have been moved from x ?

We require that $\mu(x)$ mass to be moved from X at the point x .

3. How to defined probability in the form $\int_X c(x) dF(x)$?

If we choose $c(x) = 1$ then we obtain $\int_X dF(x)$ which is the cdf of the distribution, and is therefore the probability over the region X .

4. What should the cost be to move mass from $x \in X$ to itself?

Since total variation is a norm, the norm of zero must be zero so the cost of moving mass to yourself needs to be zero.

5. How will the integral change with finite support?

The integral will become a sum divided by the size of the support. This is equivalent to assuming that the distribution integrates to one.

It is clear that if we need the cost of moving from a distribution to itself to be zero, but we want to catch a probability elsewhere, one candidate for this is the cost function:

$$C(x, y) = \mathbb{1}_{\{x \neq y\}}$$

Under this cost function, our integral over the joint distribution of the densities becomes a probability that the two random variables are not equal, which is our goal. We now wish to add constraints to the optimal transport problem that ensures that the random variables are a coupling of the two distributions.

A coupling requires that $\Pr(X = x) = \mu(x)$ and that $\Pr(Y = y) = \nu(x)$. Our transport function must meet these same requirements.

$$\int_X T(x, y) dx = \nu(y) \quad \int_Y T(x, y) dy = \mu(x)$$

Under these conditions our optimal transport problem can be written as such:

$$\inf_T \left\{ \int_{X \times Y} \mathbb{1}_{\{x \neq y\}} T(x, y) dx dy \mid \int_X T(x, y) dx = \nu(y) \quad \int_Y T(x, y) dy = \mu(x) \right\}$$

We can see that this considers all functions of the distributions (random variables) such that their marginal distributions have the same mass as the defined distributions, and minimizes the probability that they are different. This is exactly the question of finding the optimal coupling, and therefore must give the total variation norm.

5 Exercise 14

Give an example of a pair (P, π) where P samples from π but is not aperiodic.

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \pi = \left(\frac{1}{2}, \frac{1}{2} \right)$$

We may note that this matrix has an expected return time of 2 from either state, and reaches it in 2 steps with probability 1. Therefore its stationary distribution is $\pi = (.5, .5)$. All stages have period 2, so the matrix is not aperiodic.

Give an example of a pair (P, π) where P samples from π and is aperiodic but does not satisfy detailed balance.

Consider the pair:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad \pi = \left(\frac{1}{2}, 0, \frac{1}{2} \right)$$

Choosing $i = 1, j = 3$ we see that $\pi_1 P_{13} = 0$ but $P_{31} \pi_3 = \frac{1}{4}$.

6 Exercise 15

Let P be a transition probability matrix. Show that the null space of P has dimension 1, i.e., $\text{rank}(\text{null}(P - I)) = 1$

This is false. Consider $P = I_3$, then $P - I = \mathbf{0}$ and the rank of the null space of this matrix is 3.

Consider the question, show that $\text{rank}(\text{null}(P - I)) \geq 1$.

From the rank-nullity theorem, we know that the rank of the matrix plus the dimension of the null space equals the number of columns. This question is equivalent to showing that a Transition probability matrix minus the identity matrix does not have full rank.

Note that the rows of $P - I$ sum to zero. This means that the final column must be the negative sum of all the previous columns, and therefore it is linearly dependent with the previous columns. This means that $P - I$ cannot be full rank, and therefore $\text{rank}(\text{null}(P - I)) \geq 1$

7 Exercise 16

Let $P^\infty = \lim P^l$, and show that $P_i^\infty = \pi$, for $i = 1, \dots, n$

Assume that P is aperiodic and irreducible, as the limiting distribution is not well-defined for periodic matrices.

These conditions give us that the second eigenvalue will be less than one, and therefore we can apply theorem 3.

$$\|\pi - \mu_l\|_2^2 \leq C\lambda_1^l$$

Where $C \in \mathbb{R}$ and $\lambda_1 < 1$. Taking the limit of both sides implies that:

$$\lim \|\pi - \mu_l\|_2^2 \leq C \lim \lambda_1^l = 0$$

Which implies that $\|\mu_l - \pi\|_2^2 \rightarrow 0$.

$$\|\mu_0 P^l - \pi\|_2^2 \rightarrow 0 \quad \forall \mu_0$$

In particular, we may take $\mu_0 = e'_i$ where e_i is an element of the standard basis for \mathbb{R}^n , and this gives us:

$$\|P_{i,\cdot}^l - \pi\|_2^2 \rightarrow 0$$

This is the result we were looking for.

8 Exercise 17

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

First we consider the case where this matrix is not diagonalizable. This is the case when $p + q = 1$. In this case the transition probability matrix can be written as:

$$P = \begin{pmatrix} q & 1-q \\ q & 1-q \end{pmatrix}$$

$$P^2 = \begin{pmatrix} q^2 + (1-q)q & (1-q)^2 + q(1-q) \\ q^2 + (1-q)q & (1-q)^2 + q(1-q) \end{pmatrix} = \begin{pmatrix} q & 1-q \\ q & 1-q \end{pmatrix}$$

In this case, the powers of this matrix are all equal to P .

Since we know that $p + q = 1$, $1 - p - q = 0$. Therefore

$$\frac{q}{p+q} + \left(1 - \frac{q}{p+q}\right) (1 - p - q)^l = q$$

This is exactly the top right element of $P^l = P$ which is $\mu_l(0)$.

Now consider the case where this matrix is diagonalizable, i.e. $p + q \neq 1$.

The eigenvalues of this matrix are $1, 1 - p - q$ with eigenvectors of

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} -\frac{p}{q} \\ 1 \end{pmatrix}$$

The diagonalization can be written as:

$$P^k = \begin{pmatrix} 1 & -\frac{p}{q} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^k \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{-q}{p+q} & \frac{q}{p+q} \end{pmatrix}$$

Multiplying back through results in

$$P^k = \frac{1}{p+q} \begin{pmatrix} q + p(1-p-q)^k & p - p(1-p-q)^k \\ q - q(1-p-q)^k & p + q(1-p-q)^k \end{pmatrix}$$

Using $\mu_0 = (1, 0)$ we get that $\mu_k = \frac{1}{p+q}(q + p(1-p-q)^k \quad p - p(1-p-q)^k)$
Considering only the first element of this vector we reach that:

$$\mu_k(0) = \frac{q}{p+q} + \frac{p}{p+q}(1-p-q)^k = \frac{q}{p+q} + \left(1 - \frac{q}{p+q}\right)(1-p-q)^k$$

From the fact that $\pi = \frac{1}{p+q}(q, p)$ we can then find that

$$|\mu_k - \pi| = \left| \left(1 - \frac{q}{p+q}\right)(1-p-q)^k + \frac{q}{p+q} - \frac{q}{p+q} \right| = \left(1 - \frac{q}{p+q}\right) |1-p-q|^k$$