DIACONIS EXERCISE 13

1. Preliminary definitions

Let P and Q be probability distributions on a finite set Ω. The *total variation distance* between P and Q, denoted $||P - Q||$ is defined by

$$||P - Q|| = \max_{A \subseteq \Omega} |P(A) - Q(A)|$$

Note that since $P(\Omega \backslash A) - Q(\Omega \backslash A) = -(P(A) - Q(A))$, an equivalent definition is

$$||P - Q|| = \max_{A \subseteq \Omega} P(A) - Q(A).$$

This definition apparently requires us to consider all events $A \subseteq \Omega$ to find the one on which P and Q assign the most widely differing probabilities. But a moments thought shows that

$$||P - Q|| = \sum_{\substack{\omega \in \Omega \\ P(w) > Q(w)}} \big(P(\omega) - Q(\omega)\big)$$

and hence that

$$||P - Q|| = \tfrac{1}{2} \textstyle\sum_{\omega \in \Omega} \big|P(\omega) - Q(\omega)\big|. \tag{1}$$

It still seems quite remarkable to me that (1) can serve as a definition of total variation distance. This equation also shows that total variation distance is essentially the same as the ℓ_1 norm on $\mathbf{R}^N$.

2. Expectation and variance of the number of fixed points

Suppose that we shuffle a pack of n cards by choosing uniformly at random two numbers from $\{1, 2, \ldots, n\}$. If the numbers are the same, we do nothing; otherwise we swap the cards in the indicated positions. The corresponding probability distribution on the symmetric group S_n is defined by $P(1_{S_n}) = 1/n$ and $P(t) = 2/n^2$ if t is a transposition. If we perform the shuffle k times, then the probability that the cards are permuted according to the permutation $\sigma \in S_n$ is $P^{\star k}(\sigma)$. Here $P^{\star k}$ is the k*-th convolution* of P, as defined by $Q^{\star 1} = Q$ and

$$P^{\star k}(\sigma) = \sum_{\pi \in S_n} P^{\star (k-1)}(\sigma \pi^{-1}) P(\pi)$$

for each $k \geq 2$.

The first part of Exercise 13 in Diaconis' book outlines a proof that if $b > 0$ and $k = \frac{n}{2}\log n - bn$ then the total variation distance between $P^{\star k}$ and the uniform distribution U is non-negligible. Earlier, in Theorem 5, it is shown that if $c > 0$ and $k = \frac{n}{2}\log n + cn$ then

$$||P^{\star k} - U|| \leq a\mathrm{e}^{-2c}$$

for a constant $a \in \mathbf{R}$, so this result is sharp. Even the crude statement of this 'fast cut-off', that for any $\varepsilon > 0$, a shuffle with $(1/2+\varepsilon)n\log n$ steps guarantees good mixing, while the shuffle obtained by $(1/2-\varepsilon)n\log n$ steps will (with overwhelming probability) be poor, seems striking.

Let $F(\sigma)$ denote the number of fixed points of $\sigma \in S_n$. Under the uniform distribution permutations with no fixed points, i.e. derangements, occur with probability about $1/\mathrm{e}$. If k is small then it is intuitively clear that the probability distribution $P^{\star k}$ will favour permutations with relatively many fixed points, and so derangements, and other permutations with few fixed points, will be underrepresented. By making this precise we shall get a bound in Corollary 9 on the variation distance $||P^{\star k} - U||$.

We shall in fact solve the more general version of the problem where $P(1_{S_n}) = p_n$ for given $p_n \in \mathbf{R}$, and $P(t) = (1-p_n)/\binom{n}{2}$ if t is a transposition; this is required to solve the second part of the exercise. For notational convenience, let $q_n = 1 - p_n$.

Proposition 1. *If $k \in \mathbf{N}$ then*

$$\mathbf{E}_{P^{\star k}}(F) = 1 + (n-1)\Big(1 - \frac{2q_n}{n-1}\Big)^k.$$

In particular, if $p_n = 1/n$ then $q_n = 1 - 1/n$ and

$$\mathbf{E}_{P^{\star k}}(F) = 1 + (n-1)\Big(1 - \frac{2}{n}\Big)^k.$$

Proof. The n-dimensional natural representation of S_n decomposes as the sum of an irreducible subrepresentation W of dimension $n-1$ and the trivial representation. The trace of the permutation matrix representing $\sigma \in S_n$ is simply $F(\sigma)$, so we we have $\mathrm{Tr}_W(\sigma) = F(\sigma) - 1$ for all $\sigma \in S_n$.

Let

$$x = p_n + \frac{q_n}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} (ij) \in \mathbf{C}S_n$$

be the element encoding the probability distribution P. By the previous paragraph we have $\mathbf{E}_P(F-1) = \mathrm{Tr}_W(x)$. More generally, since the product in the group algebra $\mathbf{C}S_n$ corresponds to the convolution product on probability distributions, we have

$$\mathbf{E}_{P^{\star k}}(F-1) = \mathrm{Tr}_W(x^k) \tag{2}$$

It follows from Lemma 2 below that x acts as $\alpha 1_W$ on W, where

$$\alpha = p_n + \frac{(1-p_n)}{\binom{n}{2}}\binom{n}{2}\frac{\chi_W\big((12)\big)}{n-1} = p_n + q_n\frac{n-3}{n}$$
$$= 1 - \frac{2q_n}{n-1}.$$

Note that in the particular case where $p_n = 1/n$ we have $\alpha = 1 - \frac{2}{n}$. Hence x^k acts as $\big(1 - \frac{2}{n-1}(1-p_n)\big)^k 1_W$ on W and, by (2), we have

$$\mathbf{E}_{P^{\star k}}(F-1) = (n-1)\Big(1 - \frac{2q_n}{n-1}\Big)^k$$

from which the result follows immediately. □

The following lemma, which can be easily proved using Schur's Lemma, was used in the proof of Proposition 1.

Lemma 2. *Let W be an irreducible representation of S_n. If $x \in \mathbf{C}S_n$ is the sum of all elements in the conjugacy class of $\sigma \in S_n$ then x acts on W as*

$$\frac{|x^{S_n}|\chi_W(\sigma)}{\chi_W(1)}1_W.$$ □

It is easily seen from (1) that if $f : S_n \to \mathbf{R}$ is any function such that $|f(\sigma)| \leq 1$ for all $\sigma \in S_n$, then

$$||Q - R|| \geq \frac{1}{2}\big|f(\sigma)\big(Q(\sigma) - R(\sigma)\big)\big|.$$

Taking $f(\sigma) = F(\sigma)/n$ and applying Proposition 1 we get

$$||P^{\star k} - U|| \geq \frac{1}{2n}\big(\mathbf{E}_{P^{\star k}}(F) - 1\big) = \Big(1 - \frac{2q_n}{n-1}\Big)^k.$$

In particular, if $p_n = 1/n$ then since $(1 - \frac{2}{n})^n \to \mathrm{e}^{-2}$ as $n \to \infty$, it follows that n steps do not suffice to get good mixing in this case. The same result holds whenever $p_n \to 0$ as $n \to \infty$. As Diaconis remarks, to get a stronger result we need to use the variance of F.

Proposition 3. *If $k \in \mathbf{N}$ and $n \geq 4$ then*

$$\mathbf{Var}_{P^{\star k}}(F) = 1 + (n-1)\Big(1 - \frac{2q_n}{n-1}\Big)^k + \frac{n(n-3)}{2}\Big(1 - \frac{4q_n}{n}\Big)^k$$
$$\frac{(n-1)(n-2)}{2}\Big(1 - \frac{4q_n}{n-1}\Big)^k - (n-1)^2\Big(1 - \frac{2q_n}{n-1}\Big)^{2k}$$

Proof. The variance of F is the same as the variance of $F-1$ so as in Proposition 1, we may work with $F-1$. We keep the notation from this proposition. To find the variance of $F-1$ we need the expected value of $(F-1)^2$. Since $(F-1)(\sigma)^2$ is the trace of σ in its action on $W \otimes W$ we have

(3) $$\mathbf{E}_{P^{\star k}}(F-1)^2 = \mathrm{Tr}_{W\otimes W}(x).$$

To compute this trace we decompose $W \otimes W$ into its irreducible constituents. We begin by observing that

$$\begin{aligned}\chi_{W\otimes W} &= \chi_W \times \chi_W \\ &= \chi_W \times (1_{S_{n-1}}\big\uparrow^{S_n} - 1_{S_n}) \\ &= \Big(\chi_W \big\downarrow_{S_{n-1}}\Big)\big\uparrow^{S_n} - \chi_W.\end{aligned}$$

In the standard notation for irreducible characters of the symmetric group, $\chi_W = \chi^{(n-1,1)}$. Using the ordinary branching rule (see [1, Chapter 9]) we get

$$\chi_W\big\downarrow_{S_{n-1}} = \chi^{(n-1)} + \chi^{(n-2,1)}$$

and hence, provided $n \geq 4$,

$$\chi_W\big\downarrow_{S_{n-1}}\big\uparrow^{S_n} - \chi_W = \chi^{(n)} + \chi^{(n-1,1)} + \chi^{(n-2,2)} + \chi^{(n-2,1,1)}.$$

Therefore W decomposes as a direct sum of four irreducible representations. By Lemma 2, the scalar by which x acts on an irreducible representation U is

$$p_n + q_n \frac{\chi_U\big((12)\big)}{\chi_U(1)}.$$

The table below shows $\chi^\lambda(1)$ and $\chi^\lambda(12)$ for the irreducible characters appearing above. These values are easily computed using the Murnagham–Nakayama rule: see [1, Chapter 21]. The calculations can be simplified by using the identity $(n-3)(n-4)/2 = (n-2)(n-5)/2 + 1$.

λ	$\chi^\lambda(1)$	$\chi^\lambda\big((12)\big)$
$(n-1,1)$	$n-1$	$n-3$
$(n-2,2)$	$n(n-3)/2$	$(n-3)(n-4)/2$
$(n-2,1,1)$	$(n-1)(n-2)/2$	$(n-2)(n-5)/2$

It follows that there is a basis of W on which x acts as the matrix

$$I_1 \oplus \Big(1 - \frac{2q_n}{n-1}\Big) I_{n-1} \oplus \Big(1 - \frac{4q_n}{n}\Big) I_{n(n-3)/2} \oplus \Big(1 - \frac{4q_n}{n-1}\Big) I_{(n-1)(n-2)/2}.$$

Hence by (3), we have

$$\begin{aligned}\mathbf{E}_{P^{\star k}}(F-1)^2 = 1 + (n-1)\Big(1 - \frac{2q_n}{n-1}\Big)^k + \\ \frac{n(n-3)}{2}\Big(1 - \frac{4q_n}{n}\Big)^k + \frac{(n-1)(n-2)}{2}\Big(1 - \frac{4q_n}{n-1}\Big)^k.\end{aligned}$$

The proposition now follows on subtracting

$$\Big(\mathbf{E}_{P^{\star k}}(F-1)\Big)^2 = (n-1)^2\Big(1 - \frac{2q_n}{n-1}\Big)^{2k}$$

using the value given in Proposition 1. □

The following special case is worth noting.

Proposition 4. *If $p_n = 1/n$ then*

$$\mathbf{Var}_{P^{\star k}}(F) = 1 + (n-1)\Big(1 - \frac{2}{n}\Big)^k - \frac{(n+1)(n-2)}{2}\Big(1 - \frac{2}{n}\Big)^{2k} + \frac{(n-1)(n-2)}{2}\Big(1 - \frac{4}{n}\Big)^k.$$

Proof. Substituting $q_n = (n-1)/n$ in Proposition we see that

$$\frac{n(n-3)}{2}\Big(1 - \frac{4q_n}{n}\Big)^k = \frac{n(n-3)}{2}\Big(1 - \frac{2}{n}\Big)^{2k}$$

and

$$(n-1)^2\Big(1 - \frac{2q_n}{n-1}\Big)^{2k} = (n-1)^2\Big(1 - \frac{2}{n}\Big)^{2k}.$$

The difference of these expressions gives the third term above, and the others come from direct substitution. □

To get the corollary of Proposition 1 and Proposition 4 when $k = \frac{n}{2}\log n - bn$ we need the following lemma.

Lemma 5. *If $f(n)$ is a polynomial of degree d with leading term an^d and $(d - r_n)\log n \to 0$ as $n \to \infty$ then*

$$f(n)\Big(1 - \frac{2r_n}{n}\Big)^{\frac{n}{2}\log n - bn} = a\mathrm{e}^{2bd}\Big(1 + O\big(\frac{\log n}{n}\big)\Big)$$

as $n \to \infty$.

Proof. It is not hard to show that $\log f(n) = d\log n + \log a + O(1/n)$ and that

$$\begin{aligned}\log\Big(1 - \frac{2r_n}{n}\Big)^{\frac{n}{2}\log n - bn} &= \big(\frac{n}{2}\log n - bn\big)\log\Big(1 - \frac{2r_n}{n}\Big) \\ &= -\big(\frac{n}{2}\log n - bn\big)\Big(\frac{2r_n}{n} + O(1/n^2)\Big) \\ &= -r_n\log n + 2br_n + O\big(\frac{\log n}{n}\big).\end{aligned}$$

The lemma now follows from the hypothesis that $(d - r_n)\log n \to 0$ as $n \to \infty$. □

Note that the proof of the lemma makes it clear that the implied constant in the $O(1/n)$ term depends on b.

Corollary 6. *Let $k = \frac{n}{2}\log n - bn$. Then*

$$\mathbf{E}_{P^{\star k}}(F) \to 1 + \mathrm{e}^{2b},$$
$$\mathbf{Var}_{P^{\star k}}(F) \to 1 + \mathrm{e}^{2b},$$

provided that $p_n \log n \to 0$ as $n \to \infty$, *Proof.* By Lemma 5 in the case when $r_n = nq_n/(n-1)$ and $f(n) = (n-1)^2$ we get

$$\lim_{n\to\infty}(n-1)^2\Big(1-\frac{2q_n}{n-1}\Big)^k = \mathrm{e}^{2b}.$$

To apply the lemma we need that

$$\Big(1-\frac{nq_n}{n-1}\Big)\log n \to 0 \quad \text{as } n\to\infty;$$

this holds because

$$-\frac{\log n}{n} \le \Big(1-\frac{nq_n}{n-1}\Big)\log n \le p_n \log n$$

and by assumption $p_n \log n \to 0$ as $n \to \infty$. Hence, by Proposition 1 we have

$$\lim_{n\to\infty} \mathbf{E}_{P^{\star k}}(F) = 1 + \mathrm{e}^{2b}.$$

The proof of the limit for $\mathbf{Var}_{P^{\star k}}(F)$ is similar using the lemma and Proposition 2. □

3. Bounds

The following two propositions will be used to turn this corollary into a bound on $||P^{\star k} - U||$.

Proposition 7. *Suppose that* $p_n \log n \to 0$ *as* $n \to \infty$. *Let* $M \in \mathbf{N}$. *If* $k = \frac{n}{2}\log n - bn$ *where* $b \ge 2$ *and* $M = \lfloor \mathrm{e}^{2b}/2 \rfloor$ *then*

$$\mathbf{P}_{P^{\star k}}(F \le M) \le \frac{2}{M}$$

for all k *sufficiently large.*

Proof. Choose k sufficiently large so that $\mathbf{E}_{P^{\star k}}(F) \ge \mathrm{e}^{2b}$ and $\mathbf{Var}_{P^{\star k}}(F) \le 2 + e^{2b}$. Since $(1+e^b)^2 \ge 2 + e^{2b}$, it follows from Chebychev's inequality that

$$\mathbf{P}_{P^{\star k}}\Big(F \le \mathrm{e}^{2b} - t(1+\mathrm{e}^b)\Big) \le \frac{1}{t^2}.$$

Putting $t = \frac{1}{2}(\mathrm{e}^b - 1)$ we get

$$\mathbf{P}_{P^{\star k}}\Big(F \le \frac{\mathrm{e}^{2b}}{2}\Big) \le \frac{2}{(\mathrm{e}^b-1)^2}.$$

One can check that

$$(\mathrm{e}^b - 1)^2 \ge \mathrm{e}^{2b}/2$$

for all $b \ge 2$. It follows that if $M = \lfloor \mathrm{e}^{2b}/2 \rfloor$ then

$$\mathbf{P}_{P^{\star k}}(F \le M) \le \mathbf{P}_{P^{\star k}}\Big(F \le \frac{\mathrm{e}^{2b}}{2}\Big) \le \frac{2}{(\mathrm{e}^b-1)^2} \le \frac{2}{\mathrm{e}^{2b}/2} \le \frac{2}{M}$$

as required. □

Proposition 8. *Let $M \in \mathbf{N}$. Then*

$$\mathbf{P}_U(F > M) \le \frac{1}{(M+1)!}.$$

Proof. The probability that a particular $M+1$-subset of $\{1, 2, \ldots, n\}$ is fixed by a permutation in S_n is $(n-M-1)!/n!$; now sum over all $M+1$-subsets to get

$$\mathbf{P}_U(F > M) \le \binom{n}{M+1} \frac{(n-M-1)!}{n!} = \frac{1}{(M+1)!}. \qquad \square$$

Corollary 9. *Suppose that $p_n \log n \to 0$ as $n \to \infty$. If $k = \frac{n}{2}\log n - bn$ where $b \ge 2$ then*

$$||P^{\star k} - U|| \ge 1 - \frac{6}{\mathrm{e}^{2b} - 2}$$

provided k is sufficiently large.

Proof. Let $M = \lfloor \mathrm{e}^{2b}/2 \rfloor$ and consider the event $F \le M$. We have

$$||P^{\star k} - U|| \ge P_U(F \le M) - P_{P^{\star k}}(F \le M).$$

It follows from the previous two propositions that, provided k is sufficiently large

$$||P^{\star k} - U|| \ge 1 - \frac{1}{(M+1)!} - \frac{2}{M} \ge 1 - \frac{3}{M} \ge 1 - \frac{6}{\mathrm{e}^{2b} - 2}$$

as required. $\square$

This gives a non-trivial bound provided that $b \ge 1.04$. Therefore provided $p_n = O(1/\log n)$ we have shown that $\frac{n}{2}\log n - 2n$ steps do not suffice to get good mixing.

4. Second part of Exercise 13

The second part of Exercise 13 claims that if $p_n = 1/(1 + \binom{n}{2})$, so the identity is equally likely to be chosen as any transpositions, then $c(n)n^2$ are necessary to get a good shuffle, where $c(n) \to \infty$ as $n \to \infty$.

It follows easily from Proposition 1 and the inequality $1 - x \le \mathrm{e}^{-x}$ that

$$0 \le \mathbf{E}_{P^{\star c(n)n^2}}(F - 1) \le (n-1)\mathrm{e}^{-2c(n)nq_n}$$

for all $n \in \mathbf{N}$. A similar bound will hold for the variance of F. So it seems that if $k = c(n)n^2$ then F is not able to detect any significant difference between $P^{\star k}$ and U unless p_n is at least $1/\log n$; certainly the smaller value $p_n = 1/(1 + \binom{n}{2})$, as compared to $1/n$, will not cause any unexpected problems.

It is however possible to use similar ideas to get a slightly weaker result. The key observation is that if we make much fewer than $1/p_n$ steps then there is a good chance that we have never chosen the identity. In this case, the sign of the resulting permutation is given the parity of the number of steps. So if $A_n \subseteq S_n$ is the alternating group then

$$\mathbf{P}_{P^{\star k}}(A_n) \geq (1 - p_n)^k \geq 1 - kp_n$$

whenever k is even. Since $\mathbf{P}_U(A_n) = 1/2$ for all n, it follows from the definition of total variation distance that

$$||P^{\star k} - U|| \geq 1/2 - kp_n$$

whenever k is even. So in this particular case we have

$$||P^{\star k} - U|| \geq 1/2 - \theta$$

whenever k is even and $k \leq \theta n^2/2$. So $n^2/4$ steps do not suffice.

References

[1] James, G. D. *The representation theory of the symmetric groups*, vol. 682 of *Lecture Notes in Mathematics.* Springer, Berlin, 1978.