Combinatorics, Probability and Computing (2002) 11, 607–627. © 2002 Cambridge University Press
DOI: 10.1017/S0963548302005382 Printed in the United Kingdom

Information Loss in Top to Random Shuffling

DUDLEY STARK

School of Mathematical Sciences, Queen Mary,
University of London, London E1 4NS, UK
(e-mail: D.S.Stark@maths.qmul.ac.uk)

Received 2 January 2002; revised 27 June 2002

A top to random shuffle of a deck of cards is performed by taking the top card off of the deck and replacing it in a randomly chosen position of the deck. We find approximations of the relative entropy of a deck of n cards after m successive top to random shuffles. Initially the relative entropy decays linearly and for larger m it decays geometrically at a rate that alters abruptly at $m = n \log n$. It converges to an explicitly given expression when $m = \lfloor n \log n + cn \rfloor$ for a constant c.

1. Introduction

In a top to random shuffle of a deck of n cards, the top card of the deck is removed and replaced in a random position of the deck, where the choice of position is uniform over all n possibilities. The purpose of this paper is to estimate the information loss in a deck of cards due to successive independent top to random shuffles.

A shuffle of n cards may be viewed as an element of the symmetric group S_n, while a series of independent shuffles is a random walk on S_n. In representing a top to random shuffle by a permutation we use the following description taken from p. 137 of [6]. Places in the deck are numbered 1 to n from top to bottom. Initially, cards labelled 1 to n are in perfect order. After shuffling, the state of the deck is the permutation π, with $\pi(i)$ the label of the card at position i. Thus $\pi(1) = 5$ means that the card labelled 5 is on top of the deck.

Let $P_{n,m}$ be the distribution on S_n induced by m independent top to random shuffles, so that the probability distribution of one shuffle is given by

$$P_{n,1}(\pi) = \begin{cases} \frac{1}{n} & \text{if } \pi \text{ is the identity or one of the cycles } (1\,2\,3 \cdots i), \quad 1 < i \leqslant n, \\ 0 & \text{otherwise.} \end{cases}$$

Let U_n be the uniform measure on S_n, so that $U_n(\pi) = \frac{1}{n!}$ for all $\pi \in S_n$. That the stationary

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

distribution of the Markov chain corresponding to successive top to random shuffles is uniform is not hard to prove.

The top to random shuffle was originally studied in [1], where it was shown that a deck of n cards becomes random after about $n \log n$ top to random shuffles. (Unless an alternate base (which will always be 2) is given, logarithms are assumed to be natural.) In [6] explicit transition probabilities were found for the top to random shuffle and were used to find the limiting total variation distance from the uniform distribution of the distribution of a deck of n cards shuffled $\lfloor n \log n + cn \rfloor$ times, where c is a real constant. (Total variation distance is defined in Section 4.2.)

The relative entropy of $P_{n,m}$ to stationarity

$$H(P_{n,m}|U_n) = \sum_{\pi \in S_n} P_{n,m}(\pi) \log_2 \left(\frac{P_{n,m}(\pi)}{U_n(\pi)} \right)$$

is a quantity measuring the amount of information remaining in the deck of cards after m shuffles. In this paper we get asymptotic estimates for $H(P_{n,m}|U_n)$ for m fairly small as well as for m near the threshold region, where $H(P_{n,m}|U_n)$ converges.

We state our results in two theorems, the first of which is effective for fairly small values of m.

Theorem 1.1. *Let $\alpha = \log_2 e \doteq 1.4427$. If $m \leqslant n$, then the relative entropy is bounded above and below by*

$$\log_2 n! - m \log_2 n \leqslant H(P_{n,m}|U_n) \leqslant \log_2 n! - m \log_2 n + m \log_2 m + \alpha. \tag{1.1}$$

If, in addition, $m = o\left(\frac{\log_2 n}{\log_2 \log_2 n}\right)$, then

$$H(P_{n,m}|U_n) = \log_2 n! - m \log_2 n + o(\log_2 n)$$

and

$$H(P_{n,m+1}|U_n) - H(P_{n,m}|U_n) = -\log_2 n + o(\log_2 n).$$

The second result estimates $H(P_{n,m}|U_n)$ when m is close to $n \log n$. Given sequences a_n and b_n, we write $a_n \sim b_n$ as notation for $\lim_{n \to \infty} a_n / b_n = 1$.

Theorem 1.2. *Let α denote the constant $\alpha = \log_2 e$. Suppose that m satisfies*

$$n \log n - n \log(\log n - 2 \log \log n - \log 40) \leqslant m \leqslant n \log n + n \log \log n. \tag{1.2}$$

(1) *If $m = n \log n - n\omega(n)$ for some function $\omega(n)$ such that $\omega(n) \to \infty$ as $n \to \infty$, then*

$$H(P_{n,m}|U_n) \sim \alpha \omega(n) e^{\omega(n)}.$$

(2) *If $m = \lfloor n \log n + cn \rfloor$ for a real number c, then $H(P_{n,m}|U_n)$ converges and its limit is given by*

$$\lim_{n \to \infty} H(P_{n,m}|U_n) = \alpha \sum_{l=1}^{\infty} \frac{l}{(l+1)!} \Lambda(l) \log \Lambda(l),$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

where

$$\Lambda(l) = \exp\left(-e^{-c}\right) \sum_{u=0}^{l} e^{-cu}.$$

(3) *If* $m = n \log n + n\omega(n)$ *for some function* $\omega(n)$ *such that* $\omega(n) \to \infty$ *as* $n \to \infty$, *then*

$$H(P_{n,m}|U_n) \sim \frac{\alpha}{8} e^{-4\omega(n)}.$$

It is natural to consider n shuffles at a time, for then we have the following result.

Corollary 1.3. *If* m *satisfies* (1.2) *and* $m = n \log n - n\omega(n)$ *for some function* $\omega(n)$ *such that* $\omega(n) \to \infty$ *as* $n \to \infty$, *then*

$$\frac{H(P_{n,m+n}|U_n)}{H(P_{n,m}|U_n)} \to e^{-1} \quad \text{as} \quad n \to \infty.$$

If m *satisfies* (1.2) *and* $m = n \log n + n\omega(n)$ *for some function* $\omega(n)$ *such that* $\omega(n) \to \infty$ *as* $n \to \infty$, *then*

$$\frac{H(P_{n,m+n}|U_n)}{H(P_{n,m}|U_n)} \to e^{-4} \quad \text{as} \quad n \to \infty.$$

Thus, the relative entropy decreases in a geometric manner near $m = n \log n$ and the rate of decrease changes abruptly at $m = n \log n$.

Results similar to those of this paper, motivated by a numerical study in [9], were found in [8] for the Gilbert–Shannon–Reeds (GSR) model of riffle shuffling, For the GSR model, the threshold for convergence of relative entropy is $\frac{3}{2} \log_2 n + c$. Corollary 1 of [8] states that

$$H(P_{n,m+1}|U_n) - H(P_{n,m}|U_n) = -n + o(n)$$

whenever $m - \log_2 n \to -\infty$. When $m - \log_2 n \to +\infty$, the ratios of successive relative entropies satisfy

$$\frac{H(P_{n,m+1}|U_n)}{H(P_{n,m}|U_n)} = \frac{1}{4} + o(1).$$

When $m = \frac{3}{2} \log_2 n + c$, $H(P_{n,m}|U_n)$ converges to $\frac{\alpha}{24} 4^{-c}$.

Hence, for the riffle shuffle the relative entropy initially decreases linearly, as it does for the top to random shuffle. For both shuffles the relative entropy eventually decays geometrically. However, for the riffle shuffle, the rate of decay does not depend on whether m is smaller or larger than the threshold for convergence, as opposed to top to random shuffling, for which the rate changes.

Although in [8] it was possible to get good approximations of $H(P_{n,m}|U_n)$ for most m for the GSR shuffle, we can only find similarly good approximations for the top to random shuffle over a rather restricted range of m. A lower bound on the probability of having no empty boxes when m balls are placed randomly into n boxes is obtained by a Poisson approximation in Lemma 4.4. This lower bound is important in Lemma 4.6. It seems possible that a better estimate of this probability could lead to approximations of $H(P_{n,m}|U_n)$ via the methods in this paper for m below the lower restriction on m

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

in (1.2). Approximations of $H(P_{n,m}|U_n)$ could be found for m above the upper restriction in (1.2) using the methods of this paper, but fairly rapidly $H(P_{n,m}|U_n)$ becomes too small for the Poisson approximation methods used here to be effective. Improvements on our approximations for $m < n$ would require a more subtle analysis of (3.1) than we have attempted here. In comparison with the GSR shuffle, a fundamental difficulty with analysing relative entropy for the top to random shuffle seems to be that the formulas for the transition probabilities are additive rather than multiplicative in nature. (The transition probabilities are given by equation (2.4) of [8] for the GSR shuffle and by (2.4) of this paper for the top to random shuffle.)

In the rest of the paper we prove Theorems 1.1 and 1.2. Section 2 states some results from [6] and proves a lemma resulting in our main expression for $H(Q_{\mu_{n,m}}|U_n)$, equation (2.5). In Sections 3 and 4, Theorems 1.1 and 1.2 are proved, respectively. In Section 3 we use simple upper and lower bounds on (2.5) to prove Theorem 1.1. In Section 4 we use Poisson approximations of $\mu_{n,m}$ for m no farther away from $n \log n$ than $n \log \log n$ to prove Theorem 1.2. The proof of Theorem 1.2 is substantially more involved than the proof of Theorem 1.1.

2. Preliminaries

This section forms the basis of the proofs of our theorems. We first introduce definitions and results from [6].

Given a permutation $\pi \in S_n$, let $L(\pi)$ be the length of the rising sequence containing n, so that $L(\pi) = l$, provided

$$\pi^{-1}(n) > \pi^{-1}(n-1) > \cdots > \pi^{-1}(n-l+1) < \pi^{-1}(n-l).$$

For each $v \in \{0, 1, 2, \ldots, n\}$, the probability measure Q_v on S_n is defined by the probability mass function

$$Q_v(\pi) = \begin{cases} 0 & \text{if } L(\pi) < n - v, \\ \frac{(n-v)!}{n!} & \text{if } L(\pi) \geqslant n - v. \end{cases}$$

Given any probability measure μ_n on $\{0, 1, 2, \ldots, n\}$, the mixture of the Q_v with respect to $\mu_n(n - \cdot)$ is defined by

$$Q_\mu(\pi) = \sum_{v=0}^{n} \mu_n(n-v)\, Q_v(\pi). \tag{2.1}$$

The proof of Proposition 3.1 of [6] shows that

$$Q_\mu(L = l) = U_n(L = l) \sum_{u=0}^{l} \mu_n(u) u!. \tag{2.2}$$

Conditional on the event $\{L(\pi) = l\}$, the measures Q_μ and U_n are both uniform, so that

$$Q_\mu(\pi | L = l) = U_n(\pi | L = l).$$

The following lemma shows that the relative entropy between Q_μ and U_n equals the relative entropy between the induced laws. A similar property was used for total variation distance in [6].

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

Lemma 2.1. *With the definitions of Q_μ and U_n above,*

$$H(Q_\mu|U_n) = \sum_{l=1}^{n} Q_\mu(L=l)\log_2\left(\frac{Q_\mu(L=l)}{U_n(L=l)}\right).$$

Proof.

$$\begin{aligned} H(Q_\mu|U_n) &= \sum_{\pi\in S_n} Q_\mu(\pi)\log_2\left(\frac{Q_\mu(\pi)}{U_n(\pi)}\right) \\ &= \sum_{l=1}^{n}\sum_{\substack{\pi\in S_n \\ L(\pi)=l}} Q_\mu(\pi)\log_2\left(\frac{Q_\mu(L=l)}{U_n(L=l)}\times\frac{Q_\mu(\pi|L=l)}{U_n(\pi|L=l)}\right) \\ &= \sum_{l=1}^{n}\sum_{\substack{\pi\in S_n \\ L(\pi)=l}} Q_\mu(\pi)\log_2\left(\frac{Q_\mu(L=l)}{U_n(L=l)}\right) \\ &= \sum_{l=1}^{n} Q_\mu(L=l)\log_2\left(\frac{Q_\mu(L=l)}{U_n(L=l)}\right). \end{aligned}$$

□

Let $P_{n,m}$ be the probability distribution on S_n induced by m independent top to random shuffles and let $\mu_{n,m}(u)$ be the probability that u boxes are unoccupied after m balls are dropped into n boxes at random. Corollary 2.1 of [6] proves the formula

$$P_{n,m}(\pi) = Q_{\mu_{n,m}}(\pi) := \sum_{v=0}^{n} \mu_{n,m}(n-v)\,Q_v(\pi). \tag{2.3}$$

This relationship between $Q_{\mu_{n,m}}$ and U_n is given by (2.2):

$$Q_{\mu_{n,m}}(L=l) = U_n(L=l)\sum_{u=0}^{l}\mu_{n,m}(u)u!. \tag{2.4}$$

Because of (2.3), (2.4), and Lemma 2.1, we have

$$H(P_{n,m}|U_n) = H(Q_{\mu_{n,m}}|U_n) = \sum_{l=1}^{n} Q_{\mu_{n,m}}(L=l)\log_2\left(\sum_{u=0}^{l}\mu_{n,m}(u)u!\right). \tag{2.5}$$

3. Asymptotics when the number of shuffles is small

In this section we prove Theorem 1.1. Assume that $m \leqslant n$, in which case $L \geqslant n-m$, and when m balls are placed randomly into n boxes there are at least $n-m$ unoccupied boxes. Because of these considerations, (2.5) becomes

$$H(Q_{\mu_{n,m}}|U_n) = \sum_{l=n-m}^{n} Q_{\mu_{n,m}}(L=l)\log_2\left(\sum_{u=n-m}^{l}\mu_{n,m}(u)u!\right). \tag{3.1}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

First we will prove the lower bound in (1.1). Note that, for $m \leqslant n$ and all $l \geqslant n-m$,

$$\begin{aligned}\sum_{u=n-m}^{l} \mu_{n,m}(u)u! &\geqslant \mu_{n,m}(n-m)\,(n-m)! \\ &= \frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}\,(n-m)! \\ &= \frac{n!}{n^m}\end{aligned}$$

and so

$$\log_2\left(\sum_{u=n-m}^{l} \mu_{n,m}(u)u!\right) \geqslant \log_2 n! - m\log_2 n$$

for $l \in [n-m, n]$. We have the lower bound

$$\begin{aligned}H(Q_{\mu_{n,m}}|U_n) &\geqslant \sum_{l=n-m}^{n} Q_{\mu_{n,m}}(L=l)\big(\log_2 n! - m\log_2 n\big) \\ &= \big(\log_2 n! - m\log_2 n\big) \sum_{l=n-m}^{n} Q_{\mu_{n,m}}(L=l) \\ &= \log_2 n! - m\log_2 n.\end{aligned}$$

Next, we will prove the upper bound in (1.1). Given $u \in [n-m, n]$, for each $S \subset \{1,2,\ldots,n\}$ such that $|S| = n-u$, let A_S be the event

$$A_S = \{\text{all boxes in } S \text{ are occupied and no box in } \{1,2,\ldots,n\} \setminus S \text{ is occupied}\},$$

and let B_S be the event

$$B_S = \{\text{no box in } \{1,2,\ldots,n\} \setminus S \text{ is occupied}\}.$$

By elementary arguments we have

$$\begin{aligned}\mu_{n,m}(u) = \mathbb{P}\left(\bigcup_{\substack{S\subset\{1,2,\ldots,n\}\\|S|=n-u}} A_S\right) &\leqslant \sum_{\substack{S\subset\{1,2,\ldots,n\}\\|S|=n-u}} \mathbb{P}(A_S) \leqslant \sum_{\substack{S\subset\{1,2,\ldots,n\}\\|S|=n-u}} \mathbb{P}(B_S) \\ &= \binom{n}{n-u}\left(\frac{n-u}{n}\right)^m.\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{u=n-m}^{l} \mu_{n,m}(u)u! &\leqslant \sum_{u=n-m}^{l} \binom{n}{n-u}\left(\frac{n-u}{n}\right)^m u! \\ &= \frac{n!}{n^m} \sum_{u=n-m}^{l} \frac{(n-u)^m}{(n-u)!}\end{aligned}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

$$\leqslant \frac{n!}{n^m} m^m \sum_{u=n-m}^{n} \frac{1}{(n-u)!}$$
$$\leqslant \frac{n!}{n^m} m^m e$$

for all $l \in [1, n]$. Therefore,

$$\log_2 \left(\sum_{u=n-m}^{l} \mu_{n,m}(u) u! \right) \leqslant \log_2 n! - m \log_2 n + m \log_2 m + \alpha$$

for all l. From (3.1), we have

$$\begin{aligned} H(Q_{\mu_{n,m}} | U_n) &\leqslant \sum_{l=n-m}^{n} Q_{\mu_{n,m}}(L = l) \big(\log_2 n! - m \log_2 n + m \log_2 m + \alpha \big) \\ &= \big(\log_2 n! - m \log_2 n + m \log_2 m + \alpha \big) \sum_{l=n-m}^{n} Q_{\mu_{n,m}}(L = l) \\ &= \log_2 n! - m \log_2 n + m \log_2 m + \alpha, \end{aligned}$$

completing the proof of Theorem 1.1. □

4. Asymptotics when the number of shuffles is near threshold

In this section we will prove Theorem 1.2. The starting point of our analysis is again (2.5), but now in the form

$$\begin{aligned} H(P_{n,m} | U_n) &= H(Q_{\mu_{n,m}} | U_n) \\ &= \sum_{l=1}^{n} Q_{\mu_{n,m}}(L = l) \log_2 \left(\sum_{u=0}^{l} \mu_{n,m}(u) u! \right) \\ &= \alpha \sum_{l=1}^{n} Q_{\mu_{n,m}}(L = l) \log \left(\sum_{u=0}^{l} \mu_{n,m}(u) u! \right), \end{aligned} \tag{4.1}$$

where $\alpha = \log_2 e$. We assume throughout this section (which contains four subsections) that m satisfies (1.2). Section 4.1 contains some assisting Poisson approximation results. In Section 4.2 the contribution of the sum

$$\alpha \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} Q_{\mu_{n,m}}(L = l) \log \left(\sum_{u=0}^{l} \mu_{n,m}(u) u! \right) \tag{4.2}$$

to (4.1) is shown to be of order $o(n^{-k})$ for each $k > 0$. The remaining terms

$$\alpha \sum_{l=1}^{\lfloor 4\lambda \log^2 n \rfloor} Q_{\mu_{n,m}}(L = l) \log \left(\sum_{u=0}^{l} \mu_{n,m}(u) u! \right) \tag{4.3}$$

are approximated by a sum associated with a Poisson approximation in Section 4.3 and the asymptotics of the approximation are determined in Section 4.4.

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

4.1. Poisson approximations

An important tool in our analysis will be a Poisson approximation of the distribution corresponding to $\mu_{n,m}$, the number of unoccupied boxes when m balls are distributed randomly among n boxes.

Let W denote the number of unoccupied boxes when m balls are distributed randomly among n boxes. The random variable W has probability distribution function $\mu_{n,m}$. The expectation of W is given by

$$\lambda := \mathbb{E}W = n\left(1 - \frac{1}{n}\right)^m. \tag{4.4}$$

We will need some estimates on λ for m in the range (1.2). First, we show that, for such m, λ has maximal order $\log n$ and minimal order $(\log n)^{-1}$.

Lemma 4.1. *Suppose that $n \geqslant 2432$, so that $\lambda > 0$ and so that the lower bound in* (1.2) *is defined, and suppose that m satisfies* (1.2). *Under these assumptions,*

$$\frac{0.99}{\log n} \leqslant \lambda \leqslant \log n. \tag{4.5}$$

Moreover,

$$\frac{5m}{n} e^{-m/n} \leqslant \frac{1}{4} e^{-\lambda}. \tag{4.6}$$

Proof. We start with

$$\lambda \leqslant ne^{-m/n} \leqslant ne^{-\log n + \log\log n} = \log n$$

(the complicated expression in the lower bound of (1.2) is needed to prove (4.6)) and

$$\begin{aligned}
\lambda &= n\exp\left(-\frac{m}{n} - \frac{m}{2n^2} - \frac{m}{3n^3} - \cdots\right) \\
&\geqslant ne^{-m/n}\exp\left(-\frac{m}{2n^2} - \frac{m}{2n^3} - \cdots\right) \\
&= ne^{-m/n}\exp\left(-\frac{m}{2n^2}\frac{1}{1-n^{-1}}\right) \\
&= ne^{-m/n}\exp\left(-\frac{m}{2n(n-1)}\right) \\
&\geqslant ne^{-m/n}\left(1 - \frac{m}{2n(n-1)}\right) \\
&\geqslant \frac{1}{\log n}\left(1 - \frac{\log n}{n-1}\right),
\end{aligned}$$

giving (4.5).

The bounds

$$\frac{m}{n} e^{-m/n} \leqslant (\log n + \log\log n)\frac{\log n}{n} \leqslant \frac{2\log^2 n}{n} \tag{4.7}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

and

$$e^{-\lambda} \geqslant \exp\bigl(-ne^{-m/n}\bigr) \geqslant \exp\bigl(-\log n + 2\log\log n + \log 40\bigr) = \frac{40\log^2 n}{n}$$

prove (4.6). □

The distribution of W is approximately Poisson when $m = \lfloor n\log n + cn\rfloor$ for a constant c (see [7], p. 105) and this approximation was used in [6] to find limits for total variation distance for the top to random shuffle. Setting $m = \lfloor n\log n + cn\rfloor$ is appropriate for total variation distance, and we consider that case in Theorem 1.2, but we need more precise Poisson approximation results because we will examine relative entropy over a wider range of m.

The total variation distance between two general measures μ_1 and μ_2 with mutual support in a finite or countable state space S is defined to be

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \frac{1}{2}\sum_{s\in S}|\mu_1(s) - \mu_2(s)|. \tag{4.8}$$

We use μ_λ to denote the Poisson distribution with mean λ, so that $\mu_\lambda(u) = e^{-\lambda}\frac{\lambda^u}{u!}$.

Lemma 4.2. *For $n \geqslant 2432$ and m satisfying* (1.2),

$$d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda) \leqslant \frac{1}{4}e^{-\lambda}.$$

Proof. We write W as

$$W = \sum_{k=1}^{n} I_k, \tag{4.9}$$

where I_k is the indicator random variable for the event that no balls are put into the kth cell. There is a powerful technique called the 'Stein–Chen method' for finding Poisson approximations of sums of indicators such as (4.9), which we will now apply. We refer the reader to [3] for a masterful presentation of the Stein–Chen method.

The I_k in (4.9) are negatively related in the sense of Chapter 2.1 of [3], and so Corollary 2.C.2 of [3] may be applied to W. The conclusion of Corollary 2.C.2 of [3] is

$$\begin{aligned} d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda) &= \frac{1}{2}\sum_{i=0}^{\infty}|\mu_{n,m}(i) - \mu_\lambda(i)| \\ &\leqslant (1\wedge\lambda)\left(1 - \frac{\mathrm{Var}(W)}{\lambda}\right), \end{aligned}$$

and hence

$$d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda) \leqslant 1 - \frac{\mathrm{Var}(W)}{\lambda} = \frac{\lambda^2 - \mathbb{E}((W)_2)}{\lambda}, \tag{4.10}$$

where $(W)_2 = W(W-1)$.

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

The right-hand side of (4.10) is bounded by

$$\begin{aligned}
\frac{\lambda^2 - \mathbb{E}((W)_2)}{\lambda} &= \lambda^{-1}\left\{n^2\left(1-\frac{1}{n}\right)^{2m} - n(n-1)\left(1-\frac{2}{n}\right)^m\right\} \\
&= \lambda^{-1}n^2\left\{\left(1-\frac{2}{n}+\frac{1}{n^2}\right)^m - \left(1-\frac{2}{n}\right)^m\right\} + \lambda^{-1}n\left(1-\frac{2}{n}\right)^m \\
&\leqslant \lambda^{-1}n^2\left\{\frac{m}{n^2}\left(1-\frac{2}{n}+\frac{1}{n^2}\right)^{m-1}\right\} + \lambda^{-1}n\left(1-\frac{2}{n}\right)^m \qquad (4.11) \\
&= \lambda^{-1}m\left(1-\frac{2}{n}+\frac{1}{n^2}\right)^{m-1} + \lambda^{-1}n\left(1-\frac{2}{n}\right)^m \\
&\leqslant 4\lambda^{-1}m\left(1-\frac{1}{n}\right)^{2m} + \lambda^{-1}n\left(1-\frac{2}{n}\right)^m \\
&= \frac{4m}{n}\left(1-\frac{1}{n}\right)^m + \left(1-\frac{n^{-1}}{1-n^{-1}}\right)^m \\
&\leqslant \frac{4m}{n}e^{-m/n} + e^{-m/n} \\
&\leqslant \frac{5m}{n}e^{-m/n},
\end{aligned}$$

where we have made use of the Mean Value Theorem at (4.11). The upper bound (4.6) on $\frac{5m}{n}e^{-m/n}$ completes the proof. □

By analogy with (2.4), we define the measure Q_{μ_λ} on $l = 1, 2, \ldots, n$ by

$$Q_{\mu_\lambda}(l) = U_n(L = l)\sum_{u=0}^{l}\mu_\lambda(u)u!. \qquad (4.12)$$

(We write $Q_{\mu_\lambda}(l)$, not $Q_{\mu_\lambda}(L = l)$, because Q_{μ_λ} may not be a probability measure on $\{1, 2, \ldots, n\}$. It does not seem to originate from an underlying probability measure on S_n by (2.1) and (2.2) because μ_λ is not a probability measure on $\{0, 1, 2, \ldots, n\}$. A referee pointed out that if $Q_{\mu_\lambda}(l)$ were instead defined by the closely related expression $Q_{\mu_\lambda}(l) = \frac{l}{(l+1)!}\sum_{u=0}^{l}\mu_\lambda(u)u!$, then $Q_{\mu_\lambda}(l)$ would define a probability measure on $\{0, 1, 2, \ldots, n\}$. The new definition is more natural in that sense; however, when used it causes some of the lemmas below to have less natural proofs.)

Let $Q_{\mu_{n,m}|L}$ denote the distributions induced on L by the probability measure $Q_{\mu_{n,m}}$. Lemma 4.3 shows that the total variation distance bound in Lemma 4.2 translates directly into a total variation distance bound between $Q_{\mu_{n,m}|L}$ and Q_{μ_λ}.

We will need the following two observations from p. 140 of [6]. Firstly, under the uniform distribution U_n on S_n, the upper tails of L are given by

$$U_n(L \geqslant l) = \frac{1}{l!}.$$

Secondly, the probability density function of L is given by

$$U_n(L = l) = \frac{l}{(l+1)!} \quad \text{for} \quad l = 1, 2, \ldots, n-1, \qquad U_n(L = n) = \frac{1}{n!}. \qquad (4.13)$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

Lemma 4.3. *The total variation distance between $Q_{\mu_{n,m}|L}$ and Q_{μ_λ} is bounded by*

$$d_{\mathrm{TV}}(Q_{\mu_{n,m}|L}, Q_{\mu_\lambda}) \leqslant d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda).$$

Proof.

$$\begin{aligned}
d_{\mathrm{TV}}(Q_{\mu_{n,m}|L}, Q_{\mu_\lambda}) &= \frac{1}{2}\sum_{l=1}^{n} |Q_{\mu_{n,m}}(L=l) - Q_{\mu_\lambda}(l)| \\
&= \frac{1}{2}\sum_{l=1}^{n} \left| U_n(L=l)\sum_{u=0}^{l} \mu_{n,m}(u)u! - U_n(L=l)\sum_{u=0}^{l} \mu_\lambda(u)u! \right| \\
&\leqslant \frac{1}{2}\sum_{l=1}^{n} U_n(L=l)\sum_{u=0}^{l} |\mu_{n,m}(u) - \mu_\lambda(u)|u! \\
&= \frac{1}{2}\sum_{u=0}^{n} u!\left(\sum_{l=u}^{n} U_n(L=l)\right)|\mu_{n,m}(u) - \mu_\lambda(u)| \\
&= \frac{1}{2}\sum_{u=0}^{n} u!\, U_n(L \geqslant u)\,|\mu_{n,m}(u) - \mu_\lambda(u)| \\
&= \frac{1}{2}\sum_{u=0}^{n} |\mu_{n,m}(u) - \mu_\lambda(u)| \\
&\leqslant d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda).
\end{aligned}$$

□

Some known results on the upper tails of the distributions corresponding to $\mu_{n,m}$ and μ_λ will be used hereafter. A bound on the upper tails of $\mu_{n,m}$ is given by Theorem 2.R of [3], again using the fact that the I_k are negatively correlated, is

$$\mathbb{P}(W \geqslant u) \leqslant 2\mathbb{P}(Z \geqslant u) \quad \text{for} \quad u > 3\lambda,$$

where Z is a Poisson-distributed random variable with mean λ. The probability of the upper tail $\mathbb{P}(Z \geqslant u)$ is related to the point probability $\mathbb{P}(Z = u)$ by

$$\mathbb{P}(Z \geqslant u) \leqslant \left(\frac{u+1}{u+1-\lambda}\right)\mathbb{P}(Z = u) \quad \text{for} \quad u+1 > \lambda$$

(see Proposition A.2.3 of [3]). In particular, for $u \geqslant \lfloor 3\lambda \log^2 n \rfloor + 1$,

$$\begin{aligned}
\mu_{n,m}(u) = \mathbb{P}(W = u) &\leqslant \mathbb{P}(W \geqslant u) \leqslant 2\mathbb{P}(Z \geqslant u) \\
&\leqslant 2\left(\frac{u+1}{u+1-\lambda}\right)\mathbb{P}(Z = u) \leqslant 3\mathbb{P}(Z = u) = 3\mu_\lambda(u).
\end{aligned} \tag{4.14}$$

We will use a large deviations estimate for the Poisson distribution taken from [2], p. 239. It is a 'Chernoff bound', which is a trivial consequence of Markov's inequality applied to e^{tZ} (for optimized t). If Z is a Poisson-distributed random variable with mean λ, then, for $\eta > 0$,

$$\mathbb{P}(Z \geqslant \lambda(1+\eta)) \leqslant \left(e^{\eta}(1+\eta)^{-(1+\eta)}\right)^{\lambda}. \tag{4.15}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

4.2. A large deviations bound

In this section we will show that (4.2), which is the expression

$$\alpha \sum_{l=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} Q_{\mu_{n,m}}(L=l)\log\left(\sum_{u=0}^{l}\mu_{n,m}(u)u!\right),$$

is of order $o(n^{-k})$ for any integer k.

Lemma 4.4 gets useful upper and lower bounds on $Q_{\mu_{n,m}}(L=l)/U_n(L=l)$.

Lemma 4.4. *For all $1 \leqslant l \leqslant n$, we have the upper bound*

$$\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} = \sum_{u=0}^{l}\mu_{n,m}(u)u! \leqslant l!, \tag{4.16}$$

and over the range $\lfloor 3\lambda \log^2 n\rfloor < l \leqslant n$ we have

$$\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} = \sum_{u=0}^{l}\mu_{n,m}(u)u! \leqslant \lfloor 3\lambda \log^2 n\rfloor!\ + 3\sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{l} e^{-\lambda}\lambda^u. \tag{4.17}$$

For all m in the range (1.2) *and $1 \leqslant l \leqslant n$, we have*

$$\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} = \sum_{u=0}^{l}\mu_{n,m}(u)u! \geqslant \mu_{n,m}(0) \geqslant \frac{1}{2}e^{-\lambda}. \tag{4.18}$$

Proof. The upper bound (4.16) follows from (2.4) and

$$\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} = \sum_{u=0}^{l}\mu_{n,m}(u)u! \leqslant l!\sum_{u=0}^{l}\mu_{n,m}(u) \leqslant l!.$$

The other upper bound (4.17) results from

$$\begin{aligned}\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} &= \sum_{u=0}^{l}\mu_{n,m}(u)u! \\ &\leqslant \lfloor 3\lambda \log^2 n\rfloor!\sum_{u=0}^{\lfloor 3\lambda \log^2 n\rfloor}\mu_{n,m}(u) + \sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{l}\mu_{n,m}(u)u! \\ &\leqslant \lfloor 3\lambda \log^2 n\rfloor! + 3\sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{l} e^{-\lambda}\lambda^u, \end{aligned}\tag{4.19}$$

where we have used (4.14) at (4.19).

The lower bound (4.18) results from

$$\begin{aligned}\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} &\geqslant \mu_{n,m}(0) \\ &\geqslant e^{-\lambda} - |\mu_{n,m}(0) - e^{-\lambda}|\end{aligned}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

$$\geqslant e^{-\lambda} - 2\, d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda)$$
$$\geqslant e^{-\lambda} - \frac{1}{2} e^{-\lambda} = \frac{1}{2} e^{-\lambda}.$$

where we have used (2.4) and Lemma 4.2. □

Lemma 4.5. *Suppose that m lies in the range* (1.2). *For each integer $k > 0$ the sum* (4.2) *is of order*

$$\left| \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} Q_{\mu_{n,m}}(L = l) \log\left(\frac{Q_{\mu_{n,m}}(L = l)}{U_n(L = l)} \right) \right| = o(n^{-k}).$$

Proof. We will first show that

$$\sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} Q_{\mu_{n,m}}(L = l) = o(n^{-k}), \tag{4.20}$$

and then use (4.20) to show that the function

$$\sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} Q_{\mu_{n,m}}(L = l) \log\left(\frac{Q_{\mu_{n,m}}(L = l)}{U_n(L = l)} \right) \tag{4.21}$$

can be bounded above and below by functions of order $o(n^{-k})$.

Rewrite the sum in (4.20) as

$$\sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} Q_{\mu_{n,m}}(L = l) = \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} U_n(L = l)\, \frac{Q_{\mu_{n,m}}(L = l)}{U_n(L = l)}. \tag{4.22}$$

By using (4.17) one sees that (4.22) is bounded above by

$$\lfloor 3\lambda \log^2 n \rfloor ! \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} U_n(L = l) \;+\; 3 \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} U_n(L = l) \sum_{u=\lfloor 3\lambda \log^2 n \rfloor + 1}^{l} e^{-\lambda} \lambda^u. \tag{4.23}$$

The first term of (4.23) equals

$$\lfloor 3\lambda \log^2 n \rfloor ! \sum_{l=\lfloor 4\lambda \log^2 n \rfloor + 1}^{n} U_n(L = l) = \lfloor 3\lambda \log^2 n \rfloor !\, U_n(L \geqslant \lfloor 4\lambda \log^2 n \rfloor + 1)$$
$$= \frac{\lfloor 3\lambda \log^2 n \rfloor !}{(\lfloor 4\lambda \log^2 n \rfloor + 1)!}.$$

The lower bound of (4.5) shows that the first term of (4.23) is of order

$$\frac{\lfloor 3\lambda \log^2 n \rfloor !}{(\lfloor 4\lambda \log^2 n \rfloor + 1)!} \leqslant \frac{\lfloor 2.97 \log n \rfloor !}{(\lfloor 3.96 \log n \rfloor + 1)!} = o(n^{-k}).$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

The second term in (4.23) equals

$$
\begin{aligned}
&3 \sum_{l=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} U_n(L=l) \sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{l} e^{-\lambda}\lambda^u \\
&= 3 \sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{\lfloor 4\lambda \log^2 n\rfloor} \left(\sum_{l=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} U_n(L=l) \right) e^{-\lambda}\lambda^u \\
&\quad + 3 \sum_{u=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} \left(\sum_{l=u}^{n} U_n(L=l) \right) e^{-\lambda}\lambda^u \\
&= 3 \sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{\lfloor 4\lambda \log^2 n\rfloor} \frac{1}{(\lfloor 4\lambda \log^2 n\rfloor+1)!} e^{-\lambda}\lambda^u + 3 \sum_{u=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} \frac{1}{u!} e^{-\lambda}\lambda^u \\
&\leqslant 3 \sum_{u=\lfloor 3\lambda \log^2 n\rfloor+1}^{n} \frac{1}{u!} e^{-\lambda}\lambda^u \\
&\leqslant 3\,\mathbb{P}(Z \geqslant \lfloor 3\lambda \log^2 n\rfloor + 1) \\
&\leqslant 3\,\mathbb{P}(Z \geqslant 3\lambda \log^2 n) \\
&\leqslant 3\left(e^{3\log^2 n-1}(3\log^2 n)^{-3\log^2 n}\right)^{0.3/\log n} \\
&= o(n^{-k}).
\end{aligned}
$$

where the last inequality comes from (4.5) and (4.15). This completes the proof of (4.20).

The bound (4.16) produces

$$\log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right) \leqslant \log n! \leqslant n\log n$$

for all $1 \leqslant l \leqslant n$, which estimate, together with (4.20), gives the upper bound on (4.21) of order $o(n^{-k})$ by taking k sufficiently large in (4.20). Similarly, (4.5) and (4.18) imply that

$$\log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right) \geqslant -\lambda - \log 2 \geqslant -\log n - \log 2$$

for all $1 \leqslant l \leqslant n$, which lower bound, along with (4.20), gives a lower bound of the order $o(n^{-k})$ in (4.21) by taking k sufficiently large. □

4.3. Poisson approximation of the main terms

The formal relative entropy of Q_{μ_λ} with respect to U_n is defined to be

$$
\begin{aligned}
H(Q_{\mu_\lambda}|U_n) &= \sum_{l=1}^{n} Q_{\mu_\lambda}(l) \log_2\left(\sum_{u=0}^{l} \mu_\lambda(u)u!\right) \\
&= \alpha \sum_{l=1}^{n} Q_{\mu_\lambda}(l) \log\left(\sum_{u=0}^{l} \mu_\lambda(u)u!\right).
\end{aligned} \tag{4.24}
$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

In Lemma 4.6 we bound $|H(Q_{\mu_{n,m}}|U_n) - H(Q_{\mu_\lambda}|U_n)|$. In Section 4.4 it will be shown that, for m in the range (1.2), the main terms of $H(Q_{\mu_\lambda}|U_n)$ dominate $|H(Q_{\mu_{n,m}}|U_n) - H(Q_{\mu_\lambda}|U_n)|$ and therefore give a good estimate for $H(Q_{\mu_{n,m}}|U_n)$.

Lemma 4.6. *If m lies in the range* (1.2)*, then*

$$H(Q_{\mu_{n,m}}|U_n) = H(Q_{\mu_\lambda}|U_n) + O\left(\frac{m}{n}e^{-m/n}\log^3 n \cdot \log\log n\right)$$
$$= H(Q_{\mu_\lambda}|U_n) + O\left(\frac{\log^5 n \cdot \log\log n}{n}\right).$$

Proof. We begin with

$$|H(Q_{\mu_{n,m}}|U_n) - H(Q_{\mu_\lambda}|U_n)|$$
$$\leqslant \alpha \left| \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} Q_{\mu_{n,m}}(L=l)\log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right) - \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} Q_{\mu_\lambda}(l)\log\left(\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right)\right|$$
$$+ \alpha \left| \sum_{l=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} Q_{\mu_{n,m}}(L=l)\log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right)\right| \tag{4.25}$$
$$+ \alpha \left| \sum_{l=\lfloor 4\lambda \log^2 n\rfloor+1}^{n} Q_{\mu_\lambda}(l)\log\left(\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right)\right|. \tag{4.26}$$

The term (4.25) was shown to be $o(n^{-k})$ for each integer $k > 0$ in Lemma 4.5. Replacing $Q_{\mu_{n,m}}$ by Q_{μ_λ} and using $\mu_\lambda(u) = \mathbb{P}(Z = u)$ in place of (4.14) in the proofs of Lemma 4.4 and Lemma 4.5 shows that (4.26) is also of order $o(n^{-k})$ for each integer $k > 0$.

The remaining term is bounded by

$$\alpha \left| \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} Q_{\mu_{n,m}}(L=l)\log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right) - \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} Q_{\mu_\lambda}(l)\log\left(\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right)\right|$$
$$= \alpha \left| \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} U_n(L=l)\left\{\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\log\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} - \frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right\}\right|$$
$$\leqslant \alpha \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} U_n(L=l)\left|\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\log\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} - \frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right|$$
$$\leqslant \alpha \sum_{l=1}^{\lfloor 4\lambda \log^2 n\rfloor} U_n(L=l)A_{n,m}(l)\left|\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} - \frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right|, \tag{4.27}$$

where $A_{n,m}(l)$ is defined for $1 \leqslant l \leqslant \lfloor 4\lambda \log^2 n\rfloor$ by

$$A_{n,m}(l) = \left(\left|\log\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)}\right| \vee \left|\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right|\right) + 1.$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

The inequality (4.27) is shown by applying the Mean Value Theorem to the function $x \log x$.

To bound $A_{n,m}(l)$ for $1 \leqslant l \leqslant \lfloor 4\lambda \log^2 n \rfloor$, we note that Lemma 4.4 and (4.5) give

$$\begin{aligned}\left| \log \frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} \right| &\leqslant \max(\lambda + \log 2, \log(\lfloor 4\lambda \log^2 n \rfloor !)) \\ &\leqslant 4\lambda \log^2 n \cdot \log\bigl(4\lambda \log^2 n\bigr) \\ &\leqslant 4 \log^3 n \cdot \log\bigl(4 \log^3 n\bigr).\end{aligned}$$

A similar argument using μ_λ instead of $\mu_{n,m}$ gives the bound

$$\left| \log \frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} \right| \leqslant \max(\lambda, \log(\lfloor 4\lambda \log^2 n \rfloor !)) \leqslant 4 \log^3 n \cdot \log\bigl(4 \log^3 n\bigr).$$

Hence, for $1 \leqslant l \leqslant \lfloor 4\lambda \log^2 n \rfloor$,

$$A_{n,m}(l) \leqslant 4 \log^3 n \cdot \log\bigl(4 \log^3 n\bigr) + 1.$$

Now, define $B_{n,m}$ by

$$\begin{aligned}B_{n,m} &:= \max_{1 \leqslant l \leqslant \lfloor 4\lambda \log^2 n \rfloor} A_{n,m}(l) \\ &\leqslant 4 \log^3 n \cdot \log\bigl(4 \log^3 n\bigr) + 1. \qquad (4.28)\end{aligned}$$

We have

$$\begin{aligned}&\alpha \left| \sum_{l=1}^{\lfloor 4\lambda \log^2 n \rfloor} Q_{\mu_{n,m}}(L=l) \log\left(\frac{Q_{\mu_{n,m}}(L=l)}{U_n(L=l)} \right) - \sum_{l=1}^{\lfloor 4\lambda \log^2 n \rfloor} Q_{\mu_\lambda}(l) \log\left(\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} \right) \right| \\ &\leqslant \alpha B_{n,m} \sum_{l=1}^{\lfloor 4\lambda \log^2 n \rfloor} |Q_{\mu_{n,m}}(L=l) - Q_{\mu_\lambda}(l)| \\ &\leqslant 2\alpha B_{n,m}\, d_{\mathrm{TV}}(Q_{\mu_{n,m}|L}, Q_{\mu_\lambda}) \\ &\leqslant 2\alpha B_{n,m}\, d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda),\end{aligned}$$

where the last inequality comes from Lemma 4.3. The bound in the statement of the lemma now follows from the bound $d_{\mathrm{TV}}(\mu_{n,m}, \mu_\lambda) \leqslant 5\frac{m}{n} e^{-m/n}$ derived in the proof of Lemma 4.2 along with the bounds (4.7) and (4.28). □

4.4. Asymptotics of the relative entropy

The next three lemmas find the asymptotics of $H(P_{n,m}|U_n)$ when $\lambda \to \infty$, λ converges and $\lambda \to 0$, respectively. We remind the reader that $H(P_{n,m}|U_n) = H(Q_{\mu_{n,m}}|U_n)$ (see (2.3)).

Lemma 4.7. *Suppose that m lies in the range* (1.2) *and that $\lambda \to \infty$. Under these assumptions, for sufficiently large n,*

$$H(P_{n,m}|U_n) = H(Q_{\mu_{n,m}}|U_n) = \alpha\lambda \log \lambda - \alpha\lambda + O(\log \lambda). \qquad (4.29)$$

Hence, if $m = n \log n - n\omega(n)$, then

$$H(Q_{\mu_{n,m}}|U_n) \sim \alpha\omega(n) e^{\omega(n)}. \qquad (4.30)$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

Proof. Suppose that n is sufficiently large that $\lambda > 2$ (for example). Then

$$\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} = \sum_{u=0}^{l} e^{-\lambda}\lambda^u = e^{-\lambda}\lambda^l \sum_{k=0}^{l} \lambda^{-k} = e^{-\lambda}\lambda^l\big(1+O\big(\lambda^{-1}\big)\big), \tag{4.31}$$

where the $O\big(\lambda^{-1}\big)$ term in the last equality is uniform over all $l \in [1,n]$. Thus,

$$\alpha \log \frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} = -\alpha\lambda + \alpha l \log\lambda + O\big(\lambda^{-1}\big)$$

and

$$H(Q_{\mu_\lambda}|U_n) = -K_n\alpha\lambda + \alpha\log\lambda \sum_{l=1}^{n} l\, Q_{\mu_\lambda}(l) + K_n\, O\big(\lambda^{-1}\big),$$

where $K_n = \sum_{l=1}^{n} Q_{\mu_\lambda}(l)$. Definition (4.8), Lemma 4.2 and Lemma 4.3 imply that

$$|K-1| = \left|\sum_{l=1}^{n} Q_{\mu_\lambda}(l) - \sum_{l=1}^{n} Q_{\mu_{n,m}}(l)\right| \leqslant d_{\mathrm{TV}}(Q_{\mu_{n,m}}, Q_{\mu_\lambda}) \leqslant \frac{1}{4}e^{-\lambda},$$

and hence

$$H(Q_{\mu_\lambda}|U_n) = -\alpha\lambda + \alpha\log\lambda \sum_{l=1}^{n} l\, Q_{\mu_\lambda}(l) + \cdot O\big(\lambda^{-1}\big).$$

We will show that

$$\sum_{l=1}^{n} l\, Q_{\mu_\lambda}(l) = \lambda + O(1), \tag{4.32}$$

proving that

$$H(Q_{\mu_\lambda}|U_n) = \alpha\lambda\log\lambda - \alpha\lambda + O(\log\lambda).$$

Lemma 4.6 will then imply (4.29).

We start with the calculation

$$\begin{aligned}
\sum_{l=1}^{n} l\, Q_{\mu_\lambda}(l) &= \sum_{l=1}^{n} l\, U_n(L=l)\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} \\
&= \big(1+O\big(\lambda^{-1}\big)\big)\sum_{l=1}^{n} l\, U_n(L=l)e^{-\lambda}\lambda^l \qquad (4.33)\\
&= \big(1+O\big(\lambda^{-1}\big)\big)\left\{\sum_{l=1}^{n-1}\frac{l^2}{(l+1)!}e^{-\lambda}\lambda^l + \frac{n}{n!}e^{-\lambda}\lambda^n\right\}, \qquad (4.34)
\end{aligned}$$

where we have used (4.31) at (4.33) and (4.13) at (4.34). Let Z be a Poisson-distributed random variable with mean λ. Then

$$\begin{aligned}
\sum_{l=1}^{n} l\, Q_{\mu_\lambda}(l) &= \big(1+O\big(\lambda^{-1}\big)\big)\left\{\frac{1}{\lambda}\mathbb{E}Z^2 - 2 + \frac{1-e^{-\lambda}}{\lambda} - \sum_{l=n}^{\infty} l^2\frac{e^{-\lambda}}{(l+1)!}\lambda^l + \frac{e^{-\lambda}\lambda^n}{(n-1)!}\right\} \\
&= \big(1+O\big(\lambda^{-1}\big)\big)\left\{\lambda - 1 + \frac{1-e^{-\lambda}}{\lambda} - \sum_{l=n}^{\infty} l^2\frac{e^{-\lambda}}{(l+1)!}\lambda^l + \frac{e^{-\lambda}\lambda^n}{(n-1)!}\right\}. \qquad (4.35)
\end{aligned}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

The sum in (4.35) is bounded by

$$\sum_{l=n}^{\infty} l^2 \frac{e^{-\lambda}}{(l+1)!}\lambda^l \leqslant \lambda \sum_{l=n}^{\infty} \frac{e^{-\lambda}}{(l-1)!}\lambda^{l-1} = \lambda \mathbb{P}(Z \geqslant n-1)$$
$$\leqslant \lambda \left(e^{(n-1)/\lambda - 1} \left(\frac{n-1}{\lambda} \right)^{-(n-1)/\lambda} \right)^{\lambda}$$
$$= \lambda e^{n-1-\lambda} \left(\frac{n-1}{\lambda} \right)^{-(n-1)},$$

where we have used (4.15). By using (4.5), it is easily seen that the term $-(n-1)\log((n-1)/\lambda)$ dominates the logarithm of the last inequality, so that $\sum_{l=n}^{\infty} l^2 \frac{e^{-\lambda}}{(l+1)!}\lambda^l = O(1)$. The last term in (4.35) is bounded by

$$\frac{\lambda^n}{(n-1)!} \leqslant \frac{(\log n)^n}{(n-1)!} = O(1),$$

proving (4.32).

The asymptotic (4.30) now follows from (4.29) and (4.4). □

Lemma 4.8. *Suppose that m lies in the range* (1.2) *and that λ converges to a constant $\lambda_0 > 0$. Define $\Lambda(l)$ to be*

$$\Lambda(l) = \sum_{u=0}^{l} \mu_{\lambda_0}(u)u! = e^{-\lambda_0} \sum_{u=0}^{l} \lambda_0^u.$$

Under these assumptions,

$$\lim_{n\to\infty} H(P_{n,m}|U_n) = \lim_{n\to\infty} H(Q_{\mu_{n,m}}|U_n) = \alpha \sum_{l=1}^{\infty} \frac{l}{(l+1)!}\Lambda(l)\log\Lambda(l).$$

Proof. From (4.12), (4.13) and (4.24), we have

$$H(Q_{\mu_\lambda}|U_n) = \alpha \sum_{l=1}^{n} U_n(L=l) \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right) \log \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right)$$
$$= \alpha \sum_{l=1}^{n-1} \frac{l}{(l+1)!} \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right) \log \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right) \tag{4.36}$$
$$+ \frac{\alpha}{n!} \left(\sum_{u=0}^{n} \mu_\lambda(u)u! \right) \log \left(\sum_{u=0}^{n} \mu_\lambda(u)u! \right). \tag{4.37}$$

Define $\overline{\lambda}$ to be

$$\overline{\lambda} = \max\left(\sup_n \lambda, 1 \right) < \infty.$$

For every fixed positive integer l we have $\lim_{n\to\infty} \mu_\lambda(u) \to \mu_{\lambda_0}(u)$ and

$$\lim_{n\to\infty} \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right) \log \left(\sum_{u=0}^{l} \mu_\lambda(u)u! \right) = \Lambda(l)\log\Lambda(l).$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

Moreover,

$$e^{-\bar{\lambda}} \leqslant \sum_{u=0}^{l} \mu_\lambda(u)u! \leqslant \sum_{u=0}^{l} \bar{\lambda}^l \leqslant (l+1)\bar{\lambda}^l,$$

and hence

$$\left|\left(\sum_{u=0}^{l} \mu_\lambda(u)u!\right)\log\left(\sum_{u=0}^{l} \mu_\lambda(u)u!\right)\right| \leqslant (l+1)\,\bar{\lambda}^l\big(\bar{\lambda} + l\log\bar{\lambda} + \log(l+1)\big).$$

In particular, the term (4.37) is bounded by

$$\begin{aligned}\frac{\alpha}{n!}\left(\sum_{u=0}^{n} \mu_\lambda(u)u!\right)\log\left(\sum_{u=0}^{n} \mu_\lambda(u)u!\right) &\leqslant \frac{(n+1)\alpha\bar{\lambda}^n}{n!}\big(\bar{\lambda} + n\log\bar{\lambda} + \log(n+1)\big)\\ &= o(1).\end{aligned}$$

The sum

$$\sum_{l=1}^{\infty} \frac{\bar{\lambda}^l}{(l-1)!}\big(\bar{\lambda} + l\log\bar{\lambda} + \log(l+1)\big)$$

converges, and so the remaining terms (4.36) converge to the stated limit because of the dominated convergence theorem.

Now, Lemma 4.6 shows that $H(Q_{\mu_{n,m}}|U_n)$ also converges to the stated limit. □

Lemma 4.9. *Suppose that m lies in the range* (1.2) *and that* $\lambda \to 0$. *Under these assumptions, for sufficiently large n,*

$$H(P_{n,m}|U_n) = H(Q_{\mu_{n,m}}|U_n) = \frac{\alpha}{8}\lambda^4 + O\big(\lambda^5\big).$$

Hence, if $m = n\log n + n\omega(n)$, *then*

$$H(Q_{\mu_{n,m}}|U_n) \sim \frac{\alpha}{8}e^{-4\omega(n)}. \tag{4.38}$$

Proof. Suppose that n is sufficiently large that $\lambda \leqslant 1/2$ (for example). Then

$$\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} = \sum_{u=0}^{l} e^{-\lambda}\lambda^u = \begin{cases} 1 - \lambda^2/2 + \lambda^3/3 - \lambda^4/8 + O\big(\lambda^5\big) & \text{if } l = 1,\\ 1 + \lambda^2/2 - 2\lambda^3/3 + 3\lambda^4/8 + O\big(\lambda^5\big) & \text{if } l = 2,\\ 1 + \lambda^2/2 + \lambda^3/3 - 5\lambda^4/8 + O\big(\lambda^5\big) & \text{if } l = 3,\\ 1 + \lambda^2/2 + \lambda^3/3 + 3\lambda^4/8 + O\big(\lambda^5\big) & \text{if } l \geqslant 4,\end{cases}$$

where the $O\big(\lambda^5\big)$ term is uniform for all $l \geqslant 1$. Therefore, for $l \in [1, n]$,

$$\begin{aligned}Q_{\mu_\lambda}(l)\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} &= U_n(L=l)\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\\ &= \begin{cases} U_n(L=1)\{-\lambda^2/2 + \lambda^3/3 + O\big(\lambda^5\big)\} & \text{if } l = 1,\\ U_n(L=2)\{\lambda^2/2 - 2\lambda^3/3 + \lambda^4/2 + O\big(\lambda^5\big)\} & \text{if } l = 2,\\ U_n(L=3)\{\lambda^2/2 + \lambda^3/3 - \lambda^4/2 + O\big(\lambda^5\big)\} & \text{if } l = 3,\\ U_n(L=l)\{\lambda^2/2 + \lambda^3/3 + \lambda^4/2 + O\big(\lambda^5\big)\} & \text{if } l \geqslant 4.\end{cases}\end{aligned}$$

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

The sum of the $Q_{\mu_\lambda}(l)\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}$ for $l\in[1,n]$ is

$$\begin{aligned}\sum_{l=1}^{n} Q_{\mu_\lambda}(l)\log\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)} &= U_n(L=1)\{-\lambda^2/2+\lambda^3/3\}\\ &\quad + U_n(L=2)\{\lambda^2/2-2\lambda^3/3+\lambda^4/2\}\\ &\quad + U_n(L=3)\{\lambda^2/2+\lambda^3/3-\lambda^4/2\}\\ &\quad + \sum_{l=4}^{n} U_n(L=l)\{\lambda^2/2+\lambda^3/3+\lambda^4/2\}+O\big(\lambda^5\big)\\ &= U_n(L=1)\{-\lambda^2-\lambda^4/2\}+U_n(L=2)\{-\lambda^3\}+U_n(L=3)\{-\lambda^4\}\\ &\quad + \sum_{l=1}^{n} U_n(L=l)\{\lambda^2/2+\lambda^3/3+\lambda^4/2\}+O\big(\lambda^5\big)\\ &= \frac{1}{2}\{-\lambda^2-\lambda^4/2\}+\frac{1}{3}\{-\lambda^3\}+\frac{1}{8}\{-\lambda^4\}\\ &\quad + \{\lambda^2/2+\lambda^3/3+\lambda^4/2\}+O\big(\lambda^5\big)\\ &= \frac{1}{8}\lambda^4+O\big(\lambda^5\big),\end{aligned}$$

where we have used (4.13) and have assumed that $n\geqslant 4$. Therefore,

$$\begin{aligned}H(Q_{\mu_\lambda}|U_n) &= \alpha\sum_{l=1}^{n} Q_{\mu_\lambda}(l)\log\left(\frac{Q_{\mu_\lambda}(l)}{U_n(L=l)}\right)\\ &= \frac{\alpha}{8}\lambda^4+O\big(\lambda^5\big).\end{aligned}$$

The lower bound of (4.5) shows that $\lambda^5\geqslant 0.00243/\log^5 n$, which goes to 0 more slowly than the $O(\cdot)$ term in the statement of Lemma 4.6.

The asymptotic (4.38) now follows from (4.4). □

Acknowledgement

The author thanks an anonymous referee for a detailed report.

References

[1] Aldous, D. and Diaconis, P. (1986) Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348.
[2] Alon, N. and Spencer, J. (1992) *The Probabilistic Method*, Wiley Interscience.
[3] Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation*, Oxford University Press, Oxford.
[4] Bayer, D. and Diaconis, P. (1992) Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** 294–313.
[5] Diaconis, P. (1988) *Group Representations in Probability and Statistics*, IMS, Hayward, CA.
[6] Diaconis, P., Fill, J. and Pitman, J. (1992) Analysis of top to random shuffles. *Combin. Probab. Comput.* **1** 135–155.

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press

[7] Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd edn, Wiley, New York.
[8] Stark, D., Ganesh, A. and O'Connell, N. (2002) Information loss in riffle shuffling. *Combin. Probab. Comput.* **11** 79–95.
[9] Trefethen, L. N. and Trefethen, L. M. (2000) How many shuffles to randomize a deck of cards?, *Proc. Royal Soc. London A* **456** 2561–2568.

https://doi.org/10.1017/S0963548302005382 Published online by Cambridge University Press