Submitted to the Annals of Statistics
arXiv: arXiv:1307.0978

SUPPLEMENT TO "ESTIMATION IN EXPONENTIAL FAMILIES ON PERMUTATIONS"

By Sumit Mukherjee

Columbia University

1. Proofs of main results. This section carries out the proof of all results of the main paper. Subsection 1.1 gives a brief description of the notion of permutation limits, mostly adapted from [8]. Subsection 1.2 states the large deviation principle for permutations in Theorem 1.1, and uses it to prove the results of this paper. Finally, subsection 1.3 gives a proof of Theorem 1.1 using permutation limits.

1.1. *Permutation limits.* The concept of permutation limits was introduced in [8] in 2011, and was motivated from graph limit theory. For a brief exposition of the theory of graph limits refer to Lovasz [9]. The central idea in permutation limit theory is that any permutation can be thought of as a probability measures $\mathcal{M}$ on $[0,1]^2$ with uniform marginals. For any $\pi \in S_n$, define a probability measure $\mu_\pi \in \mathcal{M}$ as $d\mu_\pi := f_\pi(x,y)dxdy$, where $f_\pi(x,y) = n\mathbf{1}\{(x,y) : \pi(\lfloor nx \rfloor) = \lfloor ny \rfloor\}$ is the density of μ_π with respect to Lebesgue measure. An intuitive definition of μ_π is as follows:

Partition $[0,1]^2$ into n^2 squares of side length $1/n$, and define $f_\pi(x,y) = n$ for all (x,y) in the (i,j)-th square if $\pi(i) = j$ and 0 otherwise. As an example, the measure μ_π corresponding to the permutation $\pi = (1,3,2)$ has the density of figure 1.

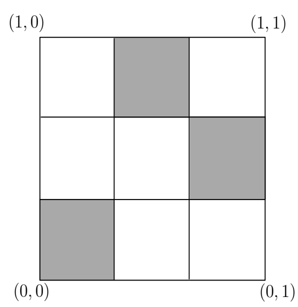

Fig 1. *Measure representation for the permutation* $(1,3,2)$. *Here the shaded region has density* 3, *and the white region has density* 0.

Here the shaded region has density 3, and the white region has 0 density.

A sequence of permutations $\pi_n \in S_n$ is said to converge to a measure $\mu \in \mathcal{M}$, if the corresponding sequence of probability measures μ_{π_n} converge weakly to μ. As an example if π_n is uniformly distributed on S_n, then π_n converges to Lebesgue measure on $[0,1]^2$. If $\pi_n = (1, 2, \cdots, n)$ is the identity permutation on S_n, then π_n converges to a measure which is uniform on the diagonal $x = y$. Similarly if $\pi_n = (n, n-1, \cdots, 1)$ is the reverse permutation, then π_n converges to the uniform measures on the diagonal $x+y = 1$. To see that non trivial limits that can arise as permutation limits, refer to Theorem 1.5 and Proposition 1.12.

1.2. *The large deviation principle.* Given a permutation π, the previous subsection defined a measure μ_π on the unit square. Also recall part (b) of theorem 1.5 which, given a permutation $\pi \in S_n$, defines a measure

$$\nu_\pi = \frac{1}{n}\sum_{i=1}^n \delta_{(i/n, \pi(i)/n)}$$

on the unit square. Both marginals of ν_π are discrete uniform on the set $\{(i/n), i \in 1, 2, \cdots, n\}$. Since the marginals are not uniform on $[0,1]$, ν_π is not an element of $\mathcal{M}$, but any weak limit of the sequence ν_{π_n} is in $\mathcal{M}$ if the size of the permutation goes to ∞. If the size of the permutation π is large, the two measure μ_π and ν_π are close in the weak topology. To see this, let $\pi \in S_n$, and let F_{μ_π} and F_{ν_π} represent the bivariate distribution functions of μ_π and ν_π respectively. Then it follows that

$$d_\infty(\mu_\pi, \nu_\pi) := \sup_{0 \le x, y \le 1} |F_{\mu_\pi}(x,y) - F_{\nu_\pi}(x,y)| \le \frac{2}{n}. \tag{1.1}$$

To see this note that both μ_π and ν_π can be defined by partitioning the unit square into n^2 boxes, such that exactly n boxes receive a mass of $1/n$. Also the choice of the n boxes is such that every row and every column will have exactly one box of positive mass. Thus any vertical line through x can intersection exactly one box in this partition which has positive probability, and so the above difference can be at most $1/n + 1/n$.

The main tool for proving the results of this paper is a large deviation principle for ν_π with respect to weak convergence on $\mathcal{M}$ where $\pi \sim \mathbb{P}_n$, the uniform probability measure on S_n. This result is stated below.

THEOREM 1.1. *If $\pi \sim \mathbb{P}_n$, the uniform measure on S_n, both the sequence of probability measures μ_π, and ν_π satisfy a large deviation principle on the* *set of probability measures on* $[0,1]^2$ *with the good rate function*

$$I(\mu) := D(\mu||u) \text{ if } \mu \in \mathcal{M}, \quad +\infty \text{ otherwise},$$

where u *is the uniform measure on* $[0,1]^2$*. More precisely, for any set* A *which is a subset of the set of probability measures on the unit square one has*

$$-\inf_{\mu\in A^o} I(\mu) \leq \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n(A) \leq \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n(A) \leq -\inf_{\mu\in\overline{A}} I(\mu),$$

where A^o *and* $\overline{A}$ *denotes the interior and closure of* A *respectively.*

Theorem 1.5 follows from Theorem 1.1 as follows.

Proof of Theorem 1.5. (a) Note that

$$e^{Z_n(f,\theta)-Z_n(0)} = \frac{1}{n!}\sum_{\pi\in S_n} e^{\theta\sum_{i=1}^n f(i/n,\pi(i)/n)} = \mathbb{E}_{\mathbb{P}_n} e^{n\theta\nu_\pi[f]},$$

where $Z_n(0) = \log n!$, and $\mu[f] = \int_{[0,1]^2} f d\mu$ denotes the mean of f with respect to μ. Since the function $\mu \mapsto \theta\mu[f]$ is bounded and continuous, an application of Varadhan's Lemma [5, Theorem 4.3.1] along with the large deviation of ν_π gives the desired conclusion.

(b) The function $\mu \mapsto \theta\mu[f] - D(\mu||u)$ is strictly concave (on the set where it is finite) and upper semi continuous on the compact set $\mathcal{M}$, and so the global maximum is attained at a unique $\mu_{f,\theta} \in \mathcal{M}$. To show the weak convergence of ν_π fix an open set U containing $\mu_{f,\theta}$, define a function $T : \mathcal{M} \mapsto [-\infty,\infty)$ by

$$T(\mu) = \theta\mu[f] \text{ if } \mu \in U^c, \quad -\infty \text{ otherwise }.$$

Then

$$\frac{1}{n}\log \mathbb{Q}_{n,f,\theta}(\nu_\pi \in U^c) = \frac{1}{n}\log \mathbb{E}_{\mathbb{P}_n} e^{nT(\nu_\pi)} - \frac{1}{n}Z_n(f,\theta).$$

Since T is upper semi continuous and bounded above, [5, Equation 4.3.2] holds trivially and so by [5, Lemma 4.3.6] along with the large deviation result for ν_π one has

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{E}_{\mathbb{P}_n} e^{nT(\nu_\pi)} \leq \sup_{\mu\in U^c\cap\mathcal{M}} \{\theta\mu[f] - D(\mu||u)\}.$$

This, along with part (a) gives

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{Q}_{n,f,\theta}(\nu_\pi\in U^c)\\ \le&\sup_{\mu\in U^c\cap\mathcal{M}}\{\theta\mu[f]-D(\mu||u)\}-\sup_{\mu\in\mathcal{M}}\{\theta\mu[f]-D(\mu||u)\}.\end{aligned}$$

The quantity on the right hand side above is negative as the infimum over the compact set $U^c\cap\mathcal{M}$ is attained, and the global minimizer $\mu_{f,\theta}$ is not in U^c by choice. This proves that $\mathbb{Q}_{n,f,\theta})(\nu_\pi\in U^c)$ decays to 0 at an exponential rate, which in particular implies that $\{\nu_\pi\}$ converges to $\mu_{f,\theta}$ weakly in probability.

(c) Since $\theta f(.)$ is integrable with respect to du, by [4, Corollary 3.2] there exists functions $a_{f,\theta}(.), b_{f,\theta}(.) :\in L^1[0,1]$ such that

$$d\mu_{a,b}=g_{a,b}dxdy:=e^{\theta f(x,y)+a_{f,\theta}(x)+b_{f,\theta}(y)}dxdy\in\mathcal{M}.$$

The proof that $\mu_{a,b}=\mu_{f,\theta}$ is by way of contradiction. Suppose this is not true. Since $\mu_{f,\theta}$ is the unique global minimizer of $I_{f,\theta}(\mu):=D(\mu||u)-\theta\mu[f]$, setting

$$h(\alpha):=I_{f,\theta}((1-\alpha)\mu_{a,b}+\alpha\mu_{f,\theta})$$

it must be that $h(\alpha)$ has a global minima at $\alpha=1$. Also

$$I_{f,\theta}(\mu_{f,\theta})\le I_{f,\theta}(u)=-\theta u(f)<\infty,$$

which forces $D(\mu_{f,\theta}||u)<\infty$. Thus letting $\phi_{f,\theta}:=\frac{d\mu_{f,\theta}}{du}$ gives

$$\begin{aligned}h'(0)=&\int_T(\phi_{f,\theta}(x,y)-g_{a,b}(x,y))(\log g_{a,b}(x,y)-\theta f(x,y))du\\ =&\int_T(\phi_{f,\theta}(x,y)-g_{a,b}(x,y))(a_{f,\theta}(x)+b_{f,\theta}(y))du\\ =&\mathbb{E}_{\mu_{f,\theta}}[a_{f,\theta}(X)+b_{f,\theta}(Y)]-\mathbb{E}_{\mu_{a,b}}[a_{f,\theta}(X)+b_{f,\theta}(Y)]=0,\end{aligned}$$

where the last equality follows from the fact that both $\mu_{f,\theta}$ and $\mu_{a,b}$ have the same uniform marginals. But h is convex, which forces that $\alpha=0$ is also a global minima of $h(.)$. Thus $h(0)=h(1)$, a contradiction to the uniqueness of $\arg\max_{\mu\in\mathcal{M}}\{\theta\mu[f]-D(\mu||u)\}$ proved in part (b). Thus it must be that

$$d\mu_{f,\theta}=d\mu_{a,b}=e^{\theta f(x,y)+a_{f,\theta}(x)+b_{f,\theta}(y)}dxdy.$$

Finally, the almost sure uniqueness of $a_{f,\theta}(.)$ and $b_{f,\theta}(.)$ follows from the uniqueness of the optimizing measure $\mu_{f,\theta}$. The last claim of part (c) then follows from part (a) by a simple calculation.

(d) Since ν_π converges in probability to $\mu_{f,\theta}$, it follows by Dominated Convergence theorem that

$$Z'_n(f,\theta) = \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} \frac{1}{n}\sum_{i=1}^n f(i/n, \pi(i)/n) \stackrel{n\to\infty}{\to} \mu_{f,\theta}[f].$$

Another application of Dominated Convergence theorem gives that

$$\frac{1}{n}[Z_n(f,\theta) - Z_n(0)] = \int_0^\theta \frac{1}{n} Z'_n(f,t)dt \stackrel{n\to\infty}{\to} \int_0^\theta \mu_{f,t}[f],$$

which along with part (a) gives that $Z'(f,\theta) = \mu_{f,\theta}[f]$.

Since $Z(f,\theta)$ is convex $Z'(f,\theta)$ is non-decreasing. To show that $Z'(f,\theta)$ is strictly increasing, by way of contradiction let $\theta_1 \neq \theta_2$ be such that $Z'(f,\theta_1) = Z'(f,\theta_2)$ for some $\theta_1 \neq \theta_2$, which implies $\mu_{f,\theta_1}[f] = \mu_{f,\theta_2}[f]$. The optimality of μ_{f,θ_1} gives

$$\theta_1 \mu_{f,\theta_1}[f] - D(\mu_{f,\theta_1}||u) \geq \theta_1 \mu_{f,\theta_2}[f] - D(\mu_{f,\theta_2}||u),$$

which implies $D(\mu_{f,\theta_1}||u) \leq D(\mu_{f,\theta_2}||u)$. By symmetry $D(\mu_{f,\theta_1}||u) = D(\mu_{f,\theta_2}||u)$, and so $\theta_1\mu_{f,\theta_1}[f] - D(\mu_{f,\theta_1}||u) = \theta_1\mu_{f,\theta_2}[f] - D(\mu_{f,\theta_2}||u)$. This implies $\mu_{f,\theta_1} = \mu_{f,\theta_2}$ by the uniqueness of theorem 1.5 part (b). By the form of the optimizer proved in theorem 1.5 part (c) one has

$$e^{\theta_1 f(x,y) + a_{f,\theta_1}(x) + b_{f,\theta_1}(y)} = e^{\theta_2 f(x,y) + a_{f,\theta_2}(x) + b_{f,\theta_2}(y)},$$

which on taking log gives

$$f(x,y) = \frac{1}{\theta_1 - \theta_2}\Big(a_{f,\theta_2}(x) + b_{f,\theta_2}(y) - a_{f,\theta_1}(x) - b_{f,\theta_1}(y)\Big).$$

Integrating with respect to y using the definition of $\mathcal{C}$ gives

$$a_{f,\theta_1}(x) - a_{f,\theta_2}(x) = \int_0^1 [b_{f,\theta_2}(y) - b_{f,\theta_1}(y)]dy,$$

and so $a_{f,\theta_1}(x) - a_{f,\theta_2}(x)$ is a constant. By symmetry $b_{f,\theta_1}(y) - b_{f,\theta_2}(y)$ is a constant as well, and so $f(x,y)$ is constant, a contradiction to the assumption that $f \in \mathcal{C}$.

Finally to show continuity of $Z'(f,\theta)$, let θ_k be a sequence of reals converging to θ. Since sequence of measures $\mu_{f,\theta_k} \in \mathcal{M}$ is tight, let μ

be any limit point of this sequence. Then by continuity of $Z(f,.)$ and lower semi continuity of $D(.||u)$ one has

$$\begin{aligned}Z(f,\theta) &= \limsup_{k\to\infty} Z(f,\theta_k)\\ &= \limsup_{k\to\infty}\{\theta_k\mu_{f,\theta_k}[f]-D(\mu_{f,\theta_k}||u)\}\le \theta\mu[f]-D(\mu||u).\end{aligned}$$

Since $Z(f,\theta)=\sup_{\mu\in\mathcal{M}}\{\theta\mu[f]-D(\mu||\theta)\}$ and the supremum is attained uniquely at $\mu_{f,\theta}$ it follows that $\mu=\mu_{f,\theta}$, and so the sequence μ_{f,θ_k} converge weakly to $\mu_{f,\theta}$. But this readily implies

$$Z'(f,\theta_k)=\mu_{f,\theta_k}[f]\stackrel{k\to\infty}{\to}\mu_{f,\theta}[f]=Z'(f,\theta),$$

and so $Z'(f,.)$ is continuous, thus completing the proof of the theorem. □

Proof of Corollary 1.7. (a) Since $\frac{1}{n}\sum_{i=1}^n f(i/n,\pi(i)/n)=\nu_\pi[f]$ and ν_f converges weakly to $\mu_{f,\theta}$ by Theorem 1.5, the desired conclusion follows.

(b) Fixing $\delta>0$ by part (a) one has

$$\begin{aligned}LD_n(\pi,\theta_0+\delta)&\stackrel{p}{\to} Z'(\theta_0)-Z'(\theta_0+\delta)<0,\\ LD_n(\pi,\theta_0-\delta)&\stackrel{p}{\to} Z'(\theta_0)-Z'(\theta_0-\delta)>0,\end{aligned}$$

and so by continuity and strict monotonicity of $Z'(f,\theta)$ from part (d) of Theorem 1.5 it follows that with probability tending to 1 there exists a unique root $\hat\theta_{LD}$ of the equation $LD_n(\pi,\theta)=0$, and $|\hat\theta_{LD}-\theta_0|\le\delta$. This proofs the consistency of $\hat\theta_{LD}$. The proof of consistency of $\hat\theta_{ML}$ follows verbatim by replacing $LD_n(\pi,\theta)$ with $ML_n(\pi,\theta)$.

(c) Since $\hat\theta_{LD}$ converges to θ_0 under $\mathbb{Q}_{n,f,\theta_0}$ and to θ_1 under $\mathbb{Q}_{n,f,\theta_1}$ the conclusion follows.

□

The following definition will be used in the proof of theorem 1.9.

Definition 1.2. *For $k\in\mathbb{N}$, partition $[0,1]^2$ into k^2 squares $\{T_{rs}\}_{r,s=1}^k$ of length $1/k$, with*

$$\begin{aligned}T_{rs} &:=\Big\{(x,y)\in T:\lceil kx\rceil=r,\lceil ky\rceil=s\Big\}\text{ for }2\le r,s,\le k,\\ T_{1s} &:=\Big\{(x,y)\in T:\lceil kx\rceil\le 1,\lceil ky\rceil=s\Big\}\text{ for }2\le s\le k,\\ T_{r1} &:=\Big\{(x,y)\in T:\lceil kx\rceil\le 1,\lceil ky\rceil=s\Big\}\text{ for }2\le r\le k,\\ T_{11} &:=\Big\{(x,y)\in T:\lceil kx\rceil\le 1,\lceil ky\rceil\le 1\Big\}.\end{aligned}$$

Also define the $k \times k$ matrix $M(\pi)$ by

$$M_{rs}(\pi) := \sum_{i=1}^{n} 1\{(i/n, \pi(i)/n) \in T_{rs}\} = n\nu_{\pi}(T_{rs}). \tag{1.2}$$

The definition ensures that T_{rs} is a disjoint partition of $[0,1]^2$, and so sum of the elements of $M(\pi)$ is n. It should be noted that all the sets T_{rs} above are μ continuity sets for any $\mu \in \mathcal{M}$. This readily follows from noting that the boundary of T_{rs} is contained in

$$\Big\{(x,y) : x = \frac{r}{k}\Big\} \cup \Big\{(x,y) : x = \frac{r-1}{k}\Big\} \cup \Big\{(x,y) : y = \frac{s}{k}\Big\} \cup \Big\{(x,y) : y = \frac{s-1}{k}\Big\},$$

which has probability 0 under any $\mu \in \mathcal{M}$, as μ has uniform marginals.

DEFINITION 1.3. *For any $k \times k$ matrix A two probability distributions $p_A, \widetilde{p}_A$ on the unit square are defined below: (Recall definition 1.8 which introduces $\mathcal{M}_k$ as the set of $k \times k$ matrices with non negative entries, such that each row and column sum equals $1/k$.)*

The measure p_A is a discrete distribution with the p.m.f. $p_A(r/k, s/k) = A_{rs}$ for $1 \le r, s \le k$.

The measure $\widetilde{p}_A$ has a density with respect to Lebesgue measure given by $p_A(x,y) =: k^2 A_{rs}$ for $x, y \in T_{rs}, 1 \le r, s \le k$. The assumption $A \in \mathcal{M}_k$ ensures that both $p_A, \widetilde{p}_A$ are probability measures, and further $\widetilde{p}_A \in \mathcal{M}$, i.e. it has uniform marginals.

PROOF OF THEOREM 1.9. (a) On applying [10, Theorem 1,2] one gets the conclusion that B_m converges to a matrix $A_{k,\theta} \in \mathcal{M}_k$ of the form $\Lambda_1 B_0 \Lambda_2$, where Λ_1 and Λ_2 are diagonal matrices.

(b) To begin note that

$$\theta \sum_{r,s=1}^{k} f(r/k, s/k) A(r,s) - 2\log k - \sum_{r,s=1}^{k} A(r,s) \log A(r,s) = \theta p_A[f] - D(p_A || p_{U_k}),$$

where $U_k \in \mathcal{M}_k$ is defined by $U_k(r,s) := \frac{1}{k^2}$. By compactness of $\mathcal{M}_k$ and strong concavity of $A \mapsto \theta p_A[f] - D(p_A || p_{U_k})$ there is a unique maximizer in $\mathcal{M}_k$, and by [4, Theorem 3.1] it follows that this maximizer is of the form $D_1 B_0 D_2$ for some diagonal matrices D_1, D_2. Since both $\Lambda_1 B_0 \Lambda_2$ and $D_1 B_0 D_2$ are in $\mathcal{M}_k$, by the uniqueness of [10, Theorem 1] one has $D_1 B_0 D_2 = \Lambda_1 B_0 \Lambda_2 = A_{k,\theta}$, thus completing the proof of part (b).

(c) Since the function $(\theta, A) \mapsto \theta p_A[f] - D(p_A||p_{U_k})$ from $\mathbb{R} \times \mathcal{M}_k$ to $[-\infty, \infty)$ is linear in θ, and has a unique maximizer $A_{k,\theta}$ in A for every θ fixed, the conclusion follows on applying Danskin's theorem [1, B.5].

(d) Since $\mu \mapsto \{\mu(T_{rs})\}_{r,s=1}^k$ is a continuous map, by theorem 1.1 and [5, Theorem 4.2.1] the matrix $\frac{1}{n}M(\pi)$ satisfies a large deviation principle on the set of $k \times k$ matrices with the good rate function

$$I_k(A) := \inf_{\mu \in \mathcal{M}: \mu(T_{rs}) = A_{rs}, 1 \le r,s \le k} D(\mu||u)$$

if $A \in \mathcal{M}_k$, and $+\infty$ otherwise. By [4, Theorem 3.1] the maximum is achieved at $\mu = \widetilde{p}_A$, and so

$$I_k(A) = D(\widetilde{p}_A||u) = \sum_{r,s=1}^k A_{rs} \log A_{rs} + 2 \log k = D(p_A||p_{U_k}).$$

An application of Varadhan's Lemma gives

$$\begin{aligned} &\frac{1}{n} \log \mathbb{E}_{\mathbb{P}_n} e^{\theta \sum_{r,s=1}^k f(r/k,s/k) M_{rs}(\pi)} \\ =&\frac{1}{n} \log \mathbb{E}_{\mathbb{P}_n} e^{n\theta \sum_{r,s=1}^k f(r/k,s/k) \nu_\pi(T_{rs})} \\ \overset{n\to\infty}{\to} & \sup_{A \in \mathcal{M}_k} \{\theta \sum_{r,s=1}^k f(r/k,s/k) A(r,s) - D(p_A||u)\} = W_k(f,\theta). \end{aligned}$$

Since

$$\begin{aligned} &|\sum_{r,s=1}^k f(r/k,s/k)\nu_\pi(T_{rs}) - \frac{1}{n}\sum_{i=1}^n f(i/n, \pi(i)/n)| \\ \le & \sup_{|x_1-x_2| \le 1/k, |y_1-y_2| \le 1/k} |f(x_1,y_1) - f(x_2,y_2)| =: \epsilon_k, \end{aligned}$$

it follows that

$$|W_k(f,\theta) - Z(f,\theta)| = \Big| \lim_{n\to\infty} \frac{1}{n} \log \frac{\mathbb{E}_{\mathbb{P}_n} e^{n\theta \sum_{r,s=1}^k f(r/k,s/k)\nu_\pi(T_{rs})}}{\mathbb{E}_{\mathbb{P}_n} e^{n\theta \sum_{i=1}^n f(i/n,\pi(i)/n)}} \Big| \le |\theta| \epsilon_k.$$

By continuity of f one has $\epsilon_k \to 0$, and so $W_k(f,\theta)$ converges to $Z(f,\theta)$.

To complete the proof assume that $p_{A_{k,\theta}} \overset{w}{\to} \mu_{f,\theta}$. In this case it follows that

$$W_k'(f,\theta) = p_{A_{k,\theta}}[f] \overset{k\to\infty}{\to} \mu_{f,\theta}[f] = Z'(f,\theta),$$

and so by Dominated Convergence we have $\lim_{k\to\infty} W_k(f,\theta) = Z(f,\theta)$. Finally since

$$\lim_{m\to\infty}\sum_{r,s=1}^{k} B_m(r,s)\log B_m(r,s) = \sum_{r,s=1}^{k} A_{k,\theta}(r,s)\log A_{k,\theta}(r,s)$$

by part (a), it follows that $Z(f,\theta)$ equals

$$\lim_{k\to\infty}\lim_{m\to\infty}\{\theta\sum_{r,s=1}^{k} f(r/k,s/k)B_m(r,s) - 2\log k - \sum_{r,s=1}^{k} B_m(r,s)\log B_m(r,s)\},$$

which is the desired conclusion.

It thus remains to show that $p_{A_{k,\theta}}$ converges weakly to $\mu_{f,\theta}$ as $k\to\infty$. Since the set of probability measures on $[0,1]^2$ is compact, the sequence $p_{A_{k,\theta}}$ is tight. If $\mu\neq\mu_{f,\theta}$ be a limit point, then by joint lower semi continuity of $D(.||.)$ one has

$$\begin{aligned}\limsup_{k\to\infty} W_k(f,\theta) =& \limsup_{k\to\infty}\{\theta p_{A_{k,\theta}}[f] - D(p_{A_{k,\theta}}||p_{U_k})\}\\ \leq& \theta\mu[f] - D(\mu||u) < Z(f,\theta).\end{aligned}$$

But this is a contradiction to the fact that $W_k(f,\theta)$ converges to $Z(f,\theta)$, and hence $p_{A_{k,\theta}}$ does indeed converges to $\mu_{f,\theta}$. This completes the proof of the theorem.

□

Before proving Theorem 1.11, a general lemma is stated which constructs $\sqrt{n}$ consistent estimates of θ in permutation models. The idea of this proof is taken from [3].

LEMMA 1.4. *Let $\mathbb{R}_{n,\theta}$ be any one parameter family on S_n, and let $G_n(\pi,\theta)$ be a function on $S_n\times\mathbb{R}$ which is differentiable in θ.*

Suppose the following two conditions hold:

(a) For every $\theta_0\in\mathbb{R}$ there exists a constant $C=C(\theta_0)$ such that

$$\mathbb{E}_{\mathbb{R}_{n\theta_0}} G_n(\pi,\theta_0)^2 \leq Cn^3 \tag{1.3}$$

(b) There exists a strictly positive continuous function $\lambda:\mathbb{R}\mapsto\mathbb{R}$ such that

$$\lim_{n\to\infty}\mathbb{R}_{n,\theta_0}(G_n'(\pi,\theta)\leq -n^2\lambda(\theta), \forall\theta\in\mathbb{R}) = 1. \tag{1.4}$$

Then the equation $G_n(\pi,\theta)=0$ has a unique root in θ. Further denoting this unique root by $\hat{\theta}_n$ one has $\sqrt{n}(\hat{\theta}_n-\theta_0)$ is $O_P(1)$ under $\mathbb{R}_{n,\theta_0}$.

Proof. Fixing a large positive real M let $\mathcal{S}_n$ be defined by

$$\mathcal{S}_n := \{\pi \in S_n : |G_n(\pi, \theta_0)| \le n^{3/2} M,\ G_n'(\pi, \theta) \le -n^2 \lambda(\theta), \theta \in \mathbb{R}\}.$$

Then for $\pi \in \mathcal{S}_n$ one has

$$G_n(\pi, \theta_0 + 1) = G_n(\pi, \theta_0) + \int_{\theta_0}^{\theta_0+1} G_n'(\pi, \theta) d\theta \le n^{3/2} M - n^2 \inf_{\theta \in [\theta_0, \theta_0+1]} \lambda(\theta),$$

which is negative for all large n. Similarly it can be shown that $G_n(\pi, \theta_0 - 1) > 0$ for $\pi \in \mathcal{S}_n$. Also note that $G_n(\pi, \theta)$ is strictly monotone on $\mathcal{S}_n$, and so by continuity of $\theta \mapsto G_n(\pi, \theta)$ there exists a unique $\hat{\theta}_n$ satisfying $G_n(\pi, \hat{\theta}_n) = 0$, and $\theta_0 - 1 < \hat{\theta}_n < \theta_0 + 1$. Finally one has

$$\begin{aligned} n^{3/2} M \ge |G_n(\pi, \theta_0)| =& |G_n(\pi, \theta_0) - G_n(\pi, \hat{\theta}_n)| \\ & \ge n^2 \Big| \int_{\hat{\theta}_n}^{\theta_0} \lambda(\theta) d\theta \Big| \ge \Big[\inf_{|\theta - \theta_0| \le 1} \lambda(\theta) \Big] |\hat{\theta}_n - \theta_0|, \end{aligned}$$

and so $\sqrt{n} |\hat{\theta}_n - \theta_0| \le KM$, where $K := [\inf_{|\theta - \theta_0| \le 1} \lambda(\theta)]^{-1} < \infty$. Thus using (1.3) and (1.4) gives

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{R}_{n,\theta_0}(|\hat{\theta}_n - \theta_0| > KM) \le & \limsup_{n\to\infty} \mathbb{R}_{n,\theta}(|G_n(\pi, \theta_0)| \ge M n^{3/2}) \\ & \le \limsup_{n\to\infty} \frac{1}{M^2 n^3} \mathbb{E}_{\mathbb{R}_{n,\theta_0}} G_n(\pi, \theta_0)^2 \le \frac{C}{M^2}. \end{aligned}$$

Since the r.h.s. above can be made arbitrarily small by choosing M large, the proof of the lemma is complete.

□

Proof of Theorem 1.11. It suffices to check the two conditions (1.3) and (1.4) of Lemma 1.4 with $\mathbb{R}_{n,\theta} = \mathbb{Q}_{n,f,\theta}$ and $G_n(\pi, \theta) = PL_n(\pi, \theta)$. For checking (1.3) an exchangeable pair is constructed.

Consider the following exchangeable pair of permutations (π, π') on S_n constructed as follows:

Pick π from $\mathbb{Q}_{n,f,\theta}$. To construct π', first pick a pair (I, J) uniformly from the set of all $\binom{n}{2}$ pairs $\{(i, j) : 1 \le i < j \le n\}$, and replace $(\pi(I), \pi(J))$ by an independent pick from the conditional distribution $(\pi(I), \pi(J) | \pi(k), k \ne I, J)$. By a simple calculation, the probabilities turn out to be the following:

(i) $(\pi'(I), \pi'(J)) = (\pi(I), \pi(J))$ with probability

$$\mathbb{Q}_{n,f,\theta}(\pi(I') = \pi(I), \pi(J') = \pi(J) | \pi(k), k \ne I, J) = \frac{1}{1 + e^{\theta y_\pi(I,J)}},$$

(ii) $(\pi'(I), \pi'(J)) = (\pi(J), \pi(I))$ with probability

$$\mathbb{Q}_{n,f,\theta}(\pi(I') = \pi(J), \pi(J') = \pi(I)|\pi(k), k \neq I, J) = \frac{e^{\theta y_\pi(I,J)}}{1 + e^{\theta y_\pi(I,J)}}.$$

Set $\pi'(i) = \pi(i)$ for all $i \neq I, J$. It can be readily checked that (π, π') is indeed an exchangeable pair. Also defining

$$W(\pi) := \sum_{i=1}^n f(i/n, \pi(i)/n), \text{ and } F(\pi, \pi') := W(\pi) - W(\pi')$$

one can check from the construction of (π, π') that

$$\mathbb{E}_{\mathbb{Q}_{n,f,\theta}}[F(\pi, \pi')|\pi] = W(\pi) - \mathbb{E}_{\mathbb{Q}_{n,f,\theta}}[W(\pi')|\pi] = \frac{1}{N_n} PL_n(\pi, \theta),$$

where $P_n(\pi, \theta)$ is as defined in the statement of the Lemma, and $N_n := \frac{n(n-1)}{2}$. Thus

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi, \theta)^2 =& N_n \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi, \theta)[\mathbb{E}_{\mathbb{Q}_{n,f,\theta}} F(\pi, \pi')|\pi] \\
=& N_n \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi, \theta) F(\pi, \pi') \\
=& N_n \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi', \theta) F(\pi', \pi) \\
=& - N_n \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi', \theta) F(\pi, \pi') \\
=& \frac{N_n}{2} \mathbb{E}_{\mathbb{Q}_{n,f,\theta}} (PL_n(\pi, \theta) - PL_n(\pi', \theta)) F(\pi, \pi')
\end{aligned}$$

where the third line uses the exchangeability of (π, π'), the fourth line uses the anti-symmetry of F, and the last line is obtained by adding the second and fourth lines together and dividing by 2. This readily implies

$$\mathbb{E}_{\mathbb{Q}_{n,f,\theta}} PL_n(\pi, \theta)^2 = \mathbb{E}_{\mathbb{Q}_{n,\theta}} V_n(\pi) \tag{1.5}$$

where $V_n(\pi) = \frac{N_n}{2} \mathbb{E}_{\mathbb{Q}_{n,f,\theta}}[(PL_n(\pi, \theta) - PL_n(\pi', \theta)) F(\pi, \pi')|\pi)$. Letting π^{ij} denote π with the elements $(\pi(i), \pi(j))$ swapped, $V_n(\pi)$ can be written as

$$V_n(\pi) = \frac{1}{2} \sum_{1 \leq i < j \leq n} \Big[PL_n(\pi, \theta) - PL_n(\pi^{ij}, \theta)\Big] \frac{y_\pi(i,j) e^{\theta y_\pi(i,j)}}{1 + e^{\theta y_\pi(i,j)}}. \tag{1.6}$$

Also setting $M := 4 \sup_{[0,1]^2} |f|$ for any (i, j) one has

$$|PL_n(\pi, \theta) - PL_n(\pi^{ij}, \theta)| \leq 4nM,$$

using the fact that $|y_\pi(i,j)| \le M$. This along with equation (1.6) gives $|V_n(\pi)| \le 4n^3M^2$, which, along with (1.5), completes the proof of (1.3) with $C = 4M^2$.

Proceeding to check (1.4) one has

$$\begin{aligned}-\frac{1}{n^2}PL_n'(\pi,\theta) =& \frac{1}{n^2}\sum_{1\le i<j\le n} y_\pi(i,j)^2 \frac{e^{\theta y_\pi(i,j)}}{1+e^{\theta y_\pi(i,j)}}\frac{1}{1+e^{\theta y_\pi(i,j)}}\\ \ge& \frac{e^{-|\theta|M}}{8n^2}\sum_{i,j=1}^n y_\pi(i,j)^2,\end{aligned}$$

where the last inequality again uses $|y_\pi(i,j)| \le M$. Since the function $g : [0,1]^4 \mapsto \mathbb{R}$ defined by

$$g((x_1,y_1),(x_2,y_2)) := \Big[f(x_1,y_1)+f(x_2,y_2)-f(x_1,y_2)-f(x_2,y_1)\Big]^2$$

is continuous, it follows that $\nu_\pi \times \nu_\pi \stackrel{w}{\to} \mu_{f,\theta_0}\times\mu_{f,\theta_0}$ in probability by part (b) of theorem 1.5. This gives

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^n y_\pi(i,j)^2\\ =&\frac{1}{n^2}\sum_{i,j=1}^n g((i/n,\pi(i)/n),(j/n,\pi(j)/n)) = (\nu_\pi\times\nu_\pi)(g)\\ \stackrel{p}{\to}&\int_{[0,1]^4}\Big[f(x_1,y_1)+f(x_2,y_2)-f(x_1,y_2)-f(x_2,y_1)\Big]^2 d\mu_{f,\theta_0}(x_1,y_1)d\mu_{f,\theta_0}(x_2,y_2)\\ =&:\alpha(\theta),\text{ say.}\end{aligned}$$

If $\alpha(\theta)=0$, then $f(x_1,y_1)+f(x_2,y_2)=f(x_1,y_2)+f(x_2,y_1)$ almost surely. On integrating with respect to x_2, y_2 and using the fact that $f\in\mathcal{C}$ gives $f(x_1,y_1)\equiv 0$, a contradiction. Thus $\alpha(\theta)>0$, and so (1.4) holds with $\lambda(\theta)=e^{-M|\theta|}\alpha(\theta)/16$. Thus both conditions of Lemma 1.4 hold, and so the conclusion follows.

□

Proof of Proposition 1.12. (a) First it will be shown that $\mu \mapsto \theta[\mu\times\mu](h)/2$ is continuous with respect to weak topology on $\mathcal{M}$. Since $\mathcal{M}$ is separable, it suffices to work with sequences, and it suffices to check the following:

$$\mu_k\in\mathcal{M},\mu_k\stackrel{w}{\to}\mu \Rightarrow (\mu_k\times\mu_k)(x_1\le x_2,y_1\le y_2)\to\mu(x_1\le x_2,y_1\le y_2)$$

But this follows from the fact that the boundary of the set $\{x_1 \le x_2, y_1 \le y_2\}$ is a subset of $\{x_1 = x_2, 0 \le y \le 1\} \cup \{0 \le x \le 1, y_1 = y_2\}$, and $\mathbb{P}(X_1 = X_2) = 0$ where X_1, X_2 are i.i.d. with distribution $U[0,1]$. Thus $\mu \mapsto \theta[\mu \times \mu](h)/2$ is continuous on $\mathcal{M} \supset \{\mu : \overline{I}(\mu) < \infty\}$.

Now, a similar computation as in the proof of Theorem 1.5 gives

$$\begin{aligned} e^{C_n(\theta)-C_n(0)} =& \frac{1}{n!}\sum_{\pi \in S_n} e^{\frac{\theta}{n}\sum_{1\le i<j\le n} h((i/n,\pi(i)/n),(j/n,\pi(j)/n))} \\ =& \mathbb{E}_{\mathbb{P}_n} e^{n\frac{\theta}{2}[\nu_\pi \times \nu_\pi](h)}. \end{aligned}$$

It then follows by an application of Varadhan's Lemma ([5, Theorem 4.3.1]) along with theorem 1.1 (on noting that the proof of Varadhan's lemma goes through as long as the function $\mu \mapsto \theta(\mu \times \mu)(h)/2$ is continuous on the set $\{\overline{I}(\mu) < \infty\}$), that

$$C(\theta) = \lim_{n\to\infty} \frac{C_n(\theta) - C_n(0)}{n} = \sup_{\mu \in \mathcal{M}} \Big\{\frac{\theta}{2}(\mu \times \mu)(h) - D(\mu||u)\Big\}.$$

The optimization problem was solved in [11] to show that there is a unique maximizer in $\mathcal{M}$, and it has the density $\rho_\theta(.,.)$ with respect to Lebesgue measure. Plugging in the formula for $u_\theta(.,.)$ gives the formula for $C(\theta)$.

(b) Since in this case the function $C(\theta)$ is convex, differentiable with a derivative which is continuous and monotone increasing, consistency of $\widetilde{\theta}_{LD}$ and $\widetilde{\theta}_{ML}$ follow from similar arguments as in Corollary 1.7.

□

1.3. *Proof of Theorem 1.1.* By (1.1) and [5, Theorem 4.2.13] μ_π and ν_π has the same large deviation, if any. Also since $\mu_\pi \in \mathcal{M}$, it is evident that the rate function can only be finite on $\mathcal{M}$. The rest of this subsection proves the large deviation for μ_π on $\mathcal{M}$ invoking [5, Theorem 4.1.11], by choosing a suitable base for the weak topology on $\mathcal{M}$.

Definition 1.5. *Let $\mathcal{M}_{k,n}$ denote the number of non negative integer valued $k \times k$ matrices with r^{th} row sum equal to $M_r := \lceil \frac{nr}{k} \rceil - \lceil \frac{n(r-1)}{k} \rceil$ and s^{th} column sum equal to $\lceil \frac{ns}{k} \rceil - \lceil \frac{n(s-1)}{k} \rceil$, i.e.*

$$\mathcal{M}_{k,n} := \Big(M \in \mathbb{N}_0^{k^2} : \sum_{s=1}^k M_{rs} = M_r, \sum_{r=1}^k M_{rs} = M_s\Big),$$

where $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$*. Note that any* $M \in \mathcal{M}_{k,n}$ *satisfies* $\sum_{r,s=1}^k M_{rs} = n$*. Recall the matrix* $M(\pi)$ *defined in* (1.2) *as the* $k \times k$ *matrix with* $M_{rs}(\pi) = n\nu_\pi[T_{rs}]$*. The following lemma shows that* $M(\pi) \in \mathcal{M}_{k,n}$ *for all* $\pi \in S_n$*, and gives the distribution of* $M(\pi)$ *when* $\pi \sim \mathbb{P}_n$*.*

Lemma 1.6. *If* π *is distributed uniformly at random on* S_n*, the distribution of* $M(\pi)$ *is given by*

$$\mathbb{P}_n(M(\pi) = M) = \frac{\left(\prod_{r=1}^k M_r!\right)^2}{n! \prod_{r,s=1}^k M_{rs}!}$$

if $M \in \mathcal{M}_{k,n}$*, and* 0 *otherwise.*

Proof. Since

$$M_{r,s}(\pi) = \sum_{i=1}^n 1\Big\{\Big\lceil \frac{ki}{n} \Big\rceil = r, \Big\lceil \frac{k\pi(i)}{n} \Big\rceil = s\Big\},$$

it follows that

$$\sum_{s=1}^k M_{r,s}(\pi) = \sum_{i=1}^n 1\Big\{\Big\lceil \frac{ki}{n} \Big\rceil = r\Big\} = M_r,$$

and so any valid configuration M is in $\mathcal{M}_{k,n}$. So fixing a particular configuration $M \in \mathcal{M}_{k,n}$, the number of possible permutations π compatible with this configuration can be computed as follows:

For the r^{th} row there are M_r choices of indices i, and that can be allocated in boxes $\{T_{r,s}\}_{s=1}^k$ in $M_r!/\prod_{s=1}^k M_{rs}!$ ways, so that box $T_{r,s}$ receives $M_{r,s}$ indices. Taking a product over r, the number of ways to distribute the indices over the boxes is

$$\frac{\prod_{r=1}^k M_r!}{\prod_{r,s=1}^k M_{rs}!}$$

Similarly, the number of ways to distribute the targets $\{\pi(i)\}$ such that box $T_{r,s}$ receives M_{rs} targets is

$$\frac{\prod_{s=1}^k M_s!}{\prod_{r,s=1}^k M_{rs}!}$$

Finally after the above distribution box $T_{r,s}$ has M_{rs} indices and M_{rs} targets, which can then be permuted freely, and so the total number of permutations compatible with any such distribution of indices and targets is

$$\prod_{r,s=1}^k M_{rs}!$$

Combining, the total number of possible permutations π satisfying $M(\pi) = M$ is given by

$$\frac{\prod_{r=1}^{k} M_r! \prod_{s=1}^{k} M_s!}{\prod_{r,s=1}^{k} M_{rs}!}$$

Since the total number of permutations in $n!$, the proof of the claim is complete. □

REMARK 1.7. *Note that in the above proposition the row and column sums of the matrix $M(\pi)$ are free of π. The distribution of $M(\pi)$ is a multivariate generalization of the hypergeometric distribution, commonly known as the Fisher-Yates distribution. This distribution arises in statistics while testing for independence in a 2-way table in the works of Diaconis-Efron ([6],[7]).*

Before proceeding the following definitions are needed. The first definition gives a base for the weak topology on $\mathcal{M}$.

DEFINITION 1.8. *For any $\mu \in \mathcal{M}$ define a matrix $P_{k,\mu} \in [0,1]^{k^2}$ by setting $P_{k,\mu}(r,s) := \mu(T_{r,s})$. Note that T_{rs} is a μ continuity set (since $\mu \in \mathcal{M}$), and so the map $\mu \mapsto P_{k,\mu}$ is continuous on $\mathcal{M}$ with respect to weak convergence.*

One can now define a base for the weak topology on $\mathcal{M}$ as follows: Fix $k \in \mathbb{N}, \epsilon > 0, \mu_0 \in \mathcal{M}$, and define the set

$$\mathcal{M}[k,\mu_0](\epsilon) := \{\mu \in \mathcal{M} : ||P_{k,\mu} - P_{k,\mu_0}||_\infty < \epsilon\},$$

where

$$||P_{k,\mu} - P_{k,\mu_0}||_\infty := \max_{1 \le r,s \le k} |P_{k,\mu}(r,s) - P_{k,\mu_0}(r,s)|.$$

Since $\mu \mapsto P_{k,\mu}$ is continuous, the set $\mathcal{M}[k,\mu_0](\epsilon)$ is open in $\mathcal{M}$. Also, for any $\mu \in \mathcal{M}$ one has $P_{k,\mu} \in \mathcal{M}_k$. Thus the operation $A \mapsto p_A$ introduced in defintiion 1.3 maps a matrix to a probability measure, and the operation $\mu \mapsto P_{k,\mu}$ maps a probability measure to a matrix.

PROPOSITION 1.9. *The collection*

$$\mathcal{M}_0 := \{\mathcal{M}[k,\mu_0](\epsilon) : k \in \mathbb{N}; \epsilon > 0, \mu_0 \in \mathcal{M}\}$$

is a base for the weak convergence on $\mathcal{M}$.

Proof. One needs to verify that given any μ_0 and an open set U containing μ_0, there is an element U_0 from this collection $\mathcal{M}_0$ such that $\mu_0 \in U_0 \subset U$. If not, then in particular the set $\mathcal{M}[k,\mu_0](1/k)$ is not contained in U for any k, and so there exists $\mu_k \in \mathcal{M}[k,\mu_0](1/k) \cap U^c$. Then for any function f which is continuous on the unit square, one has

$$\begin{aligned}|\mu[f]-\mu_k[f]| \le &\max_{[0,1]^2}|f|||P_{k,\mu}-P_{k,\mu_0}||_\infty \\ &+2\sup_{|x_1-x_2|,|y_1-y_2|\le 1/k}|f(x_1,y_1)-f(x_2,y_2)|,\end{aligned}$$

which goes to 0 as k goes to ∞. Thus μ_k converges weakly to μ, and since U is open, one has that $\mu_k \in U$ for all large k. This is a contradiction to the assumption that $\mu_k \notin U$, and so completes the proof. □

This reduces the analysis of measures to the analysis of $k \times k$ matrices for a large but fixed k.

Definition 1.10. *For $\mu_0 \in \mathcal{M}$ define a set $\mathcal{V}[k,\mu_0](\epsilon) \subset \mathcal{M}_k$ as*

$$\mathcal{V}[k,\mu_0](\epsilon) := \{A \in \mathcal{M}_k : ||A-P_{k,\mu_0}||_\infty < \epsilon\}.$$

Since $M(\pi) \in \mathcal{M}_{k,n}$ is an integer valued matrix, all configurations in $\mathcal{V}[k,\mu_0](\epsilon)$ cannot be attained by setting $A = M(\pi)/n$. Define $\mathcal{V}_n[k,\mu_0](\epsilon)$ to be the set of all $M \in \mathcal{M}_{k,n}$ such that $\frac{1}{n}M \in \mathcal{V}[k,\mu_0]$. More precisely, $\mathcal{V}_n[k,\mu_0](\epsilon)$ is defined by

$$\mathcal{V}_n[k,\mu_0](\epsilon) := \mathcal{M}_{k,n} \cap n\mathcal{V}[k,\mu_0](\epsilon) = \Big\{M \in \mathcal{M}_{k,n} : ||\frac{1}{n}M - P_{k,\mu_0}||_\infty < \epsilon\Big\}.$$

The following lemma gives an estimate of the probability that $M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon)$.

Lemma 1.11.

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n\Big(M(\pi)\in\mathcal{V}_n[k,\mu_0](\epsilon)\Big) = -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}),$$

where p_A is as in definition 1.3.

Proof. For the proof, first assume that

$$\lim_{n\to\infty}\min_{M\in\mathcal{V}_n[k,\mu_0](\epsilon)} D(p_{M/n}||p_{U_k}) = \inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}), \tag{1.7}$$

where the definition of p_A is extended to matrices A whose row/column sums need not equal $1/k$, to accomodate for the fact that for any $M \in \mathcal{V}_n[k,\mu_0](\epsilon)$ the matrix $\frac{1}{n}M$ will not satisfy this exactly. The proof of (1.7) is deferred till the end of the lemma.

For the lower bound, note that

$$\begin{aligned}\mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon)\Big) \geq & \max_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} \mathbb{P}_n(M(\pi) = M) \\ = & \max_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} \frac{\Big(\prod_{r=1}^k M_r\Big)^2}{n! \prod_{r,s=1}^k M_{rs}!}\end{aligned}$$

where the second step uses Lemma 1.6. Now, Stirling's formula gives that there exists $C < \infty$ such that

$$\begin{aligned}|\log n! - n\log n + n| =& 0 \text{ if } n = 0 \\ =& 1 \text{ if } n = 1 \\ \leq& C\log n \text{ if } n \geq 2,\end{aligned}$$

and so

$$\frac{1}{n}\log \mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon)\Big) \geq -\min_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} D(p_{M/n}||p_{U_k}) - \frac{C_k \log n}{n}$$

for some constant $C_k < \infty$. On taking limits using (1.7) completes the proof of the lower bound.

For the upper bound note that

$$\begin{aligned}\mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon)\Big) \leq & \binom{n+k^2-1}{k^2-1} \max_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} \mathbb{P}_n(M(\pi) = M) \\ \leq & (n+k^2)^{k^2} \max_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} \mathbb{P}_n(M(\pi) = M),\end{aligned}$$

since any valid configuration M is a non negative integral solution of the equation $\sum_{r,s=1}^k M_{rs} = n$. Thus proceeding as before it follows that

$$\frac{1}{n}\log \mathbb{P}_n(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon)) \leq -\min_{M \in \mathcal{V}_n[k,\mu_0](\epsilon)} D(p_{M/n}||p_{U_k}) + \frac{C_k' \log n}{n}$$

for some other $C_k' < \infty$, which on taking limits using (1.7) completes the proof of the upper bound.

It thus remains to prove (1.7). To this effect, let $M^{(n)}$ denote the minimizing configuration on the l.h.s. of (1.7). Then $\frac{1}{n}M^{(n)}$ is a sequence in the compact set

$$\{A \subset [0,1]^{k^2} : \min_{1\le r,s\le k} A_{rs} \ge 0 : \sum_{r,s=1}^{k} A_{rs} = 1\},$$

and any convergent subsequence converges to a point in $\overline{\mathcal{V}[k,\mu_0](\epsilon)}$. Thus

$$\begin{aligned}\liminf_{n\to\infty} \min_{M\in\mathcal{V}_n[k,\mu_0](\epsilon)} D(p_{M/n}||p_{U_k}) &\ge \inf_{A\in\overline{\mathcal{V}[k,\mu_0](\epsilon)}} D(p_A||p_{U_k})\\ &= \inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}),\end{aligned}$$

where the last equality follows from since $A\mapsto D(p_A||p_{U_k})$ is continuous, completing the proof of the lower bound in (1.7).

Proceeding to prove the upper bound, it suffices to prove that for any $A\in\mathcal{V}[k,\mu_0](\epsilon)$ there exists a sequence $M^{(n)}\in\mathcal{V}_n[k,\mu_0](\epsilon)$ such that $\frac{1}{n}M^{(n)}$ converges to A as $n\to\infty$. To this effect, let $\mu\in\mathcal{M}$ be such that $P_{k,\mu}=A$. (It is easy to check that such a μ always exists for any $A\in\mathcal{M}_k$). By [8, Lemma 4.2] and [8, Lemma 5.3] there exists a sequence of permutations $\{\sigma_n\}_{n\ge1}$ with $\sigma_n\in S_n$ such that ν_{σ_n} converges weakly to μ, and so setting $M^{(n)}=M(\sigma_n)$ one has that $M^{(n)}\in\mathcal{M}_{k,n}$ and $\frac{1}{n}M^{(n)}\to P_{k,\mu}=A$. Also the set

$$W_k := \{B\in[0,1]^{k^2} : ||B-P_{k,\mu_0}||_\infty < \epsilon\}$$

is open, and since $A\in W_k$, it follows that $\frac{1}{n}M^{(n)}\in W_k$ for all large n. Since $\mathcal{V}_n[k,\mu_0](\epsilon)=nW_k\cap\mathcal{M}_{k,n}$, the proof of (1.7) is complete. □

The next and final lemma derives another technical estimate using Lemma 1.11. This lemma will be used to prove Theorem 1.1.

Lemma 1.12. *For any set $\mathcal{M}[k,\mu_0](\epsilon)$ one has*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n\Big(\mu_\pi\in\mathcal{M}[k,\mu_0](\epsilon)\Big) = -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}).$$

Proof. First note that

$$||P_{k,\mu_\pi}-\frac{1}{n}M(\pi)||_\infty = ||P_{k,\mu_\pi}-P_{k,\nu_\pi}||_\infty \le \frac{4}{n}.$$

Indeed, since each square T_{rs} has four boundaries each of which intersect in exactly one row/column of the $n\times n$ partition of the unit square, the

two quantities above can differ only if there is an element on one of these rows/columns. Since each such square has probability $1/n$ under μ_π, the maximum difference can be at most $4/n$.
Thus for any $\delta \in (0,\epsilon)$ and all n large enough,

$$\mathbb{P}_n\Big(\mu_\pi \in \mathcal{M}[k,\mu_0](\epsilon)\Big) \geq \mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon-\delta)\Big)$$

Using Lemma 1.11 gives

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon-\delta)\Big) \geq -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon-\delta)} D(p_A||p_{U_k}).$$

Letting $\delta \downarrow 0$ gives

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n\Big(\mu_\pi \in \mathcal{M}[k,\mu_0](\epsilon)\Big) \geq -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}). \tag{1.8}$$

A similar argument gives

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n\Big(M(\pi) \in \mathcal{V}_n[k,\mu_0](\epsilon+\delta)\Big) \leq -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon+\delta)} D(p_A||p_{U_k}),$$

from which, letting $\delta \downarrow 0$ gives

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n\Big(\mu_\pi \in \mathcal{M}[k,\mu_0](\epsilon)\Big) \leq -\inf_{A\in\overline{\mathcal{V}[k,\mu_0](\epsilon)}} D(p_A||p_{U_k}). \tag{1.9}$$

Combining (1.8) and (1.9) gives

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}_n\Big(\mu_\pi \in \mathcal{M}[k,\mu_0](\epsilon)\Big) = -\inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}),$$

using the continuity of $A \mapsto D(p_A||p_{U_k})$. This completes the proof of the lemma. □

Proof of Theorem 1.1. Since $\mathcal{M}_0$ is a base for the weak topology on $\mathcal{M}$, by Lemma 1.12 and [5, Theorem 4.1.11] it follows that $\mathbb{P}_n$ follows a weak large deviation principle with the rate function

$$I(\mu) = \sup_{\mathcal{M}[k,\mu_0](\epsilon)\ni\mu} \inf_{A\in\mathcal{V}[k,\mu_0](\epsilon)} D(p_A||p_{U_k}).$$

Also since $\mathcal{M}$ is compact it follows that full large deviation principle holds with the good rate function $I(.)$. It thus remains to prove that $I(\mu) = D(\mu||u)$. To this effect, first note that $\mu \in \mathcal{M}[k,\mu](1/k)$, and so

$$I(\mu) \geq \liminf_{k\to\infty} \inf_{A\in\overline{\mathcal{V}[k,\mu](1/k)}} D(p_A||p_{U_k}) = \liminf_{k\to\infty} D(p_{A_k}||p_{U_k}),$$

where A_k denotes any minimizer of $A \mapsto D(p_A||p_{U_k})$ over $\overline{\mathcal{V}[k,\mu](1/k)}$. But then p_{A_k} converges weakly to μ as $k \to \infty$. The lower semi continuity of $D(.||.)$ then implies $I(\mu) \geq D(\mu||u)$, proving the lower bound.

For the upper bound note that the first supremum is over all $\mathcal{M}[k,\mu_0](\epsilon)$ containing μ , and so with $A = P_{k,\mu} \in \mathcal{V}[k,\mu](\epsilon)$ one has

$$I(\mu) \leq \sup_{k\geq 1} D(p_{P_{k,\mu}}||p_{U_k})$$

Also note that

$$D(\mu||u) = \sup_{f\in B[0,1]^2} \Big\{ \int_{[0,1]^2} f d\mu - \log \int_{[0,1]^2} e^f du \Big\},$$

$$D(p_{P_{k,\mu}}||p_{U_k}) = \sup_{f\in B_k[0,1]^2} \Big\{ \int_{[0,1]^2} f d\mu - \log \int_{[0,1]^2} e^f du \Big\},$$

where $B[0,1]^2$ denotes the set of all bounded measurable functions on $[0,1]^2$, and $B_k[0,1]^2$ denotes the subset of $B[0,1]^2$ which is constant on every $T_{rs}, 1 \leq r,s \leq k$. Indeed, both the results follows from [5, Lemma 6.2.13]. Consequently $\sup_{k\geq 1} D(p_{P_{k,\mu}}||p_{U_k}) \leq D(\mu||u)$, thus completing the proof of the upper bound.

□

2. Statement and Proof of Proposition 2.2. This section states and proves Proposition 2.2 which characterizes the joint limiting distribution of $\{\pi(1),\cdots,\pi(n)\}$, when π is generated either from the two distributions considered in this paper. Stating the proposition requires the following definition:

DEFINITION 2.1. *For a permutation $\pi \in S_n$ define the function $\pi_n : (0,1] \mapsto (0,1]$ by setting $\pi_n(t) := \frac{\pi(\lceil nt \rceil)}{n}$.*

For any measure $\mu \in \mathcal{M}$ define a stochastic process $Z_\mu : (0,1] \mapsto (0,1]$ via the following finite dimensional distributions:

For any $k \in \mathbb{N}$ and $0 < t_1 < t_2 \cdots < t_k \leq 1$ the random variables $\{Z_\mu(t_1),\cdots,Z_\mu(t_k)\}$ are mutually independent, with $Z_\mu(t_i)$ having the law $\mathcal{L}(Y|X=t_i)$, with $(X_i,Y_i)_{1\leq i\leq k} \overset{i.i.d.}{\sim} \mu$.

PROPOSITION 2.2. *(a) If π is an observation from $\mathbb{Q}_{n,f,\theta}$ as in (1.1) with $f \in \mathcal{C}$, then for any $k \in \mathbb{N}$ and $0 < t_1 < \cdots < t_k \leq 1$ one has*

$$\{\pi_n(t_1),\cdots,\pi_n(t_k)\} \overset{d}{\to} \{Z_{\mu_{f,\theta}}(t_1),\cdots,Z_{\mu_{f,\theta}}(t_k)\},$$

where $\mu_{f,\theta}$ is as in theorem 1.5.

(b) *If π is an observation from the Mallows model $M_{n,\theta}$ with Kendall's Tau as in Proposition 1.12, then for any $k \in \mathbb{N}$ and $0 < t_1 < \cdots < t_k \le 1$ one has*

$$\{\pi_n(t_1), \cdots, \pi_n(t_k)\} \stackrel{d}{\to} \{Z_{\mu_{\rho_\theta}}(t_1), \cdots, Z_{\mu_{\rho_\theta}}(t_k)\},$$

where μ_{ρ_θ} is the measure induced by the density ρ_θ of proposition 1.12.

PROOF. (a) By theorem 1.5 part (b) one has π converges to $\mu_{f,\theta}$ in probability. The result then follows by an application [2, Prop 6.1] along with [2, Cor 6.4].

(b) By [11, Theorem 1] one has that π converges to μ_{ρ_θ} in probability. The result then follows by an application [2, Prop 6.1] along with [2, Lemma 7.1].

□

As an application of the above proposition, it immediately follows for example

$$\frac{1}{n}\pi(\lceil nt\rceil) \stackrel{d}{\to} Z_\mu(t) \sim \mathcal{L}(Y_1|X_1 = t)$$
$$\mathbb{P}(\pi(\lceil nt\rceil) > \pi(\lceil ns\rceil)) \stackrel{n\to\infty}{\to} \mathbb{P}(Z_\mu(t) > Z_\mu(s)) = \mathbb{P}(Y_1 > Y_2|X_1 = t, X_2 = s),$$

where $(X_i, Y_i)_{i=1,2} \stackrel{i.i.d.}{\sim} \mu$, where μ is the relevant limiting measure. An explicit evaluation of these quantities for the models of the form $\mathbb{Q}_{n,f,\theta}$ requires the computation of the limit $\mu_{f,\theta}$.

References.

[1] D. Bertsekas, *Nonlinear Programming*, Athena Scientific Publishing, Belmont, MA, 1999.

[2] B. Bhattacharya and S. Mukherjee, Degree Sequence of Random Permutation Graphs, Available at http://arxiv.org/abs/1503.03582, 2015.

[3] S. Chatterjee. Estimation in spin glasses: A first step. *The Annals of Statistics*, 35(5):1931–1946, 2007.

[4] I. Csisźar. **I**-Divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[5] A. Dembo and O. Zeitouni. *Large deviations techniques and applications (second edition)*, *Application of Mathematics* (38), Springer, 1998.

[6] P. Diaconis and B. Efron, Testing for independence in a two-way table: New interpretations of the Chi-square statistic, *The Annals of Statistics*, Vol. 13 (3), 845–913, 1985.

[7] P. Diaconis and B. Efron, Probabilistic-geometric theorems arising from the analysis of contingency tables, *Contributions to the Theory and Application of Statistics, A Volume in Honor of Herbert Solomon* , Academic Press, 103–125, 1987.

[8] C. Hoppen, Y. Kohayakawa, C. Moreira, B. Rath, and I. Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory Series B*, 103(1):93–113, 2013.

[9] L. Lovász. *Large networks and graph limits*(60), AMS, 2012.

[10] R. Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, 35(2):876-879, (1964).

[11] S. Starr. Thermodynamic limit for the Mallows model on S_n. *Journal of Mathematical Physics*, 50(9), 2009.

[12] J. Trashorras. Large deviations for symmetrised empirical measures. *Journal of Theoretical Probability*, 21(2):397–412, 2008.

1255 Amsterdam Avenue,
New York, NY-10027.
E-mail: sm3949@columbia.edu