ESTIMATION OF PARAMETERS IN NON UNIFORM MODELS ON PERMUTATIONS

SUMIT MUKHERJEE

Abstract. Using large deviation results for a uniformly random permutation in $S_n$, asymptotics of normalizing constants are computed and limits of permutations obtained for some non uniform distributions on $S_n$. A pseudo-likelihood type estimator is shown to be consistent in a class of one parameter exponential families on permutations. A new proof of the large deviation principle of a uniformly random permutation on $S_n$ is given.

1. Introduction

Let $S_n$ denote the set of all permutations of $[n] := \{1, 2, \ldots, n\}$. An important class of non-uniform models on $S_n$ are known as Mallows models, which are of the form

$$e^{-(\theta/n)d(\pi, \pi_0) - Z_n(\theta, \pi_0)},$$

where $d(.,.) : S_n \times S_n \mapsto [0, \infty)$ is a metric on permutations, and $e^{Z_n(\theta, \pi_0)}$ is the appropriate normalizing constant. In this model $\pi_0 \in S_n$ is a location parameter, and $\theta \in \mathbb{R}$ is a shape parameter which determines the fluctuations of the model about $\pi_0$. By definition a metric $d(.,.)$ satisfies the following conditions:

$$d(\pi, \sigma) \geq 0, \text{ with equality iff } \pi = \sigma,$$

$$d(\pi, \sigma) = d(\sigma, \pi),$$

$$d(\pi, \sigma) \leq d(\pi, \tau) + d(\sigma, \tau).$$

Another restriction which is usually required on $d(.,.)$ is that $d(.,.)$ is right invariant, i.e.

$$d(\pi, \sigma) = d(\pi \tau, \sigma \tau), \text{ for all } \pi, \sigma, \tau \in S_n.$$

The justification for this is that $\tau$ can be thought of as an arbitrary relabeling of the original objects $\{1, 2, \ldots, n\}$, and the distance between two permutations should be invariant to any such labeling ([D, Chapter 6]). All distances $d(.,.)$ considered in this paper are right invariant.

Some of the common choices of right invariant metric $d(.,.)$ are the following:

(a) $\sum_{i=1}^{n}|\pi(i) - \sigma(i)|$ (Foot rule)

(b) $\sum_{i=1}^{n}(\pi(i) - \sigma(i))^2$ (Spearman’s rank correlation)

(c) $\#\{1 \leq i \leq n : \pi(i) \neq \sigma(i)\}$ (Hamming Distance)

Date: May 7, 2014.

2010 Mathematics Subject Classification. 60F10, 05A05, 62F12.

Key words and phrases. Permutation, Large Deviation, Mallows Model, Pseudolikelihood.
(d) Minimum number of pairwise adjacent transpositions which converts $\pi$ into $\sigma$ (Kendall’s Tau)
(e) Minimum number of adjacent transpositions which converts $\pi$ into $\sigma$ (Cayley’s distance)

(f) $n$ – Length of the longest increasing subsequence in $\sigma \pi^{-1}$

See [D, Ch-5,6] for more details on these models.

Note that if $d(.,.)$ is right invariant, then the normalizing constant is free of $\pi_0$, as

$$\sum_{\pi \in S_n} e^{-(\theta/n)d(\pi,\pi_0)} = \sum_{\pi \in S_n} e^{-(\theta/n)d(\pi \pi_0^{-1},e)} = \sum_{\sigma \in S_n} e^{-(\theta/n)d(\sigma,e)},$$

where $e$ is the identity permutation. For the remainder of this paper the log normalizing constant will be denoted by $Z_n(\theta)$ instead of $Z_n(\theta,\pi_0)$. In all these models $\theta = 0$ corresponds to the uniform distribution, and the model shows no attraction nor repulsion towards $\pi_0$. On the other hand if $\theta$ is large, then a random pick from this model is close to $\pi_0$ in the metric $d(.,.)$.

In some of these models, using special structure of the specific model, $Z_n(\theta)$ can be computed analytically, either exactly or in the limit. See [DR, (2.9)] for an explicit answer for Kendall’s Tau, and [SS] for an asymptotic answer for the same model. However in general the normalizing constant $Z_n(\theta)$ might not be analytically computable. Also numerical computations are infeasible, as $|S_n| = n!$ grows very fast. Thus computation of MLE for $\theta$ becomes a hard problem, even if $\pi_0$ is assumed to be known.

1.1. Viewing permutations as measures. In order to define limits of permutations it is necessary to introduce a common topological space containing permutations of all orders. With this pre-requisite in mind consider the following definition, first introduced in [HKMRS].

Definition 1.1. Given a permutation $\pi \in S_n$, define the probability measure $L(\pi)$ on the unit square, henceforth denoted by $T$, given by the following density $\phi_\pi(.,.)$ with respect to the Lebesgue measure on $T$:

$$\phi_\pi(x,y) = n1\{(x,y) : \pi(\lfloor nx \rfloor) = \lfloor ny \rfloor\}.$$ 

$L(\pi)$ has the following interpretation:

Partition the unit square $T$ into $n^2$ squares of length $1/n$, producing $n$ squares in each row and column. For $i \in [n]$, if $\pi(i) = j$, then the value of the density $\phi_\pi(.,.)$ is $n$ on the $(i,j)^{th}$ unit square. Thus each row and each column has exactly one square where the density is strictly positive and equals $n$. It can be checked that

$$\int_0^1 \phi_\pi(x,z)dz = 1 \text{ and } \int_0^1 \phi_\pi(z,y)dz = 1 \text{ for every } x,y \in [0,1],$$

and so $L(\pi)$ is a probability measure on $T$ with uniform marginals.

Denote by $\mathcal{P}(T)$ the set of all probability measures on $T$, and denote by $\mathcal{U}$ the subset of $\mathcal{P}(T)$ with both marginals uniform. Then $L$ defines a 1-1 map from $S := \cup_{n=1}^\infty S_n$ into $\mathcal{U} \subset \mathcal{P}(T)$. For $\pi \in S$, the size of $\pi$ is defined to be $n$ if $\pi \in S_n$. 

Thus for every $\pi \in S$ this defines a representative probability measures in $\mathcal{P}(T)$, which will be taken as the underlying topological space. The topology equipped on $\mathcal{P}(T)$ is the topology of weak convergence, introduced in the following definition:

**Definition 1.2.** Let $\{\mu_k\}_{k=1}^\infty, \mu \in \mathcal{P}(T)$. Then $\mu_k \xrightarrow{w} \mu$ if $\mu_k(f) \converges \mu(f)$ for all continuous functions $f$ on $T$, where $\mu(f) := \int f d\mu$.

The weak topology is compatible with the bounded Lipschitz metric $d_L(\mu, \nu)$ defined as

$$d_L(\mu, \nu) := \sup_{f \in \mathcal{C}(T) : |f|_{L(T)} \leq 1} |\mu(f) - \nu(f)|,$$

where

$$\mathcal{L}(f) := \sup_{(x, y) \in T} |f(x, y)| + \sup_{(x_1, y_1), (x_2, y_2) \in T} \frac{|f(x_1, y_1) - f(x_2, y_2)|}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}},$$

and $\mathcal{C}(T)$ is the space of all continuous functions on $T$.

For more details on this metric see [Dudley, Page 394]. The following definition formally introduces the notion of convergence of permutations.

**Definition 1.3.** A sequence of permutations $\pi_k \in S$ is said to converge to a probability measure $\mu \in \mathcal{U}$, if the corresponding probability measures $L(\pi_k)$ converges weakly to $\mu$.

As an example of such convergence, if $\pi_k$ is the identity permutation on $S_k$, then $\pi_k$ converges to the measure which is uniform on the diagonal $\{(x, y) \in T : x = y\}$. If $\pi_k \in S_k$ is the permutation which sends $i$ to $k + 1 - i$ for every $i$, then $\pi_k$ converges to the uniform distribution on the other diagonal $\{(x, y) \in T : x + y = 1\}$. If $\pi_k$ is a permutation chosen uniformly at random from $S_k$, then $\pi_k \xrightarrow{w} u$, where $u$ is the Lebesgue measure on $T$.

This notion of convergence of permutations was introduced and developed in [HKMRS]. It was shown in [HKMRS, Claim 2.4] that if the size of a permutation sequence $|\pi_k| \xrightarrow{w} \infty$, then a convergent sequence of permutations is eventually constant. Recall that size of a permutation in $S_n$ is defined to be $n$. It also follows from [HKMRS, Lemma 4.2] and [HKMRS, Lemma 5.3] that any measure $\mu \in \mathcal{U}$ can arise as a limit of a permutation sequence $\pi_n$ with $|\pi_n| = n$.

This notion of convergence of permutations has similarities with the notion of convergence of dense graphs in the cut metric. One connection is that a graph converges in the sense of the cut metric iff all the sub graph densities converge. [HKMRS] show that that there exists permutation statistics analogous to subgraph densities (the number of inversions is an example) which determine the convergence of permutations in the sense defined above. For more on this connection, as well as the definition of dense graph convergence, see [HKMRS] and the references therein.

Another connection between graph convergence and permutation convergence is through permutation graphs. Given a permutation $\pi \in S_n$, consider a simple undirected graph $G(\pi)$ with vertices labelled $\{1, 2, \cdots, n\}$ as follows:

There is an edge between vertices $i$ and $j$ if $(\pi(i), \pi(j))$ is an inversion.

Thus the graph $G(\pi)$ encodes the inversion structure of the permutation $\pi$. It was observed in [GGKK, Page 8] that if $\pi_k$ converges to $\pi$ in the sense defined above, then the corresponding
sequence of graphs $G(\pi_k)$ converge in the sense of cut metric. There the authors also characterize
the graph limit for some particular choices of permutation sequences.

1.2. Main results of this paper. Let $\mathbb{P}_n$ denote the uniform distribution on the space of all
permutations $S_n$. This paper uses a large deviation principle for $\mathbb{P}_n$ to derive the limit of $Z_n(\theta)$
for $n$ large, for a class of non uniform distributions that include the cases (a), (b) and (d) above.
It also describes the limit of a random permutation from some non uniform models in a suitable
topology. This gives some answers to the question: “how does a random permutation from such
a model look like for large $n$?” For example, using these tools one can find out the asymptotic
number of inversions in a random pick from a non uniform model.

The first main result of this paper is the following theorem which computes the limiting value
of $\frac{1}{n}(Z_n(\theta) - Z_n(0))$ for a class of probability models on $S_n$ in terms of an optimization problem.
In particular, this theorem covers the foot rule and Spearman’s rank correlation ((a) and (b) of
the original list). Since $Z_n(0) = \log n!$, this gives an approximation for $Z_n(\theta)$. Stating the theorem
requires defining the Kullback Leibler divergence $D(\mu || \nu)$ between two probability measures, defined
by
$$D(\mu || \nu) := \int \log \frac{d\mu}{d\nu} d\mu \text{ if } \mu << \nu,$$
$$= \infty \text{ otherwise.}$$

**Theorem 1.4.** Let $f \in C(T)$, the set of all continuous functions on $T$, and consider the probability
model
$$Q_{n,f,\theta}(\pi) = e^{\theta \sum_{i=1}^{n} f(i/n, \pi(i)/n) - Z_n(f, \theta)},$$
where $Z_n(f, \theta)$ is the normalizing constant. Then

(a) \[\lim_{n \to \infty} \frac{Z_n(f, \theta) - Z_n(0)}{n} = \sup_{\mu \in \mathcal{U}} \{\theta \mu(f) - D(\mu || u)\},\]
where $u$ is the uniform distribution on $T$.

(b) Under this model $\pi_n$ converges weakly in probability to $\mu_{f,\theta}$, where $\mu_{f,\theta} \in \mathcal{U}$ the unique
maximizer of part (a).

The optimization in part (b) is over an infinite dimensional space. The following theorem gives
an implicit form for the limiting measure $\mu_{f,\theta}$ of Theorem 1.4, and also gives an iterative algorithm
to compute the density corresponding to the limiting measure $\mu_{f,\theta}$. Intuitively the algorithm starts
with the function $e^{\theta f(x,y)}$ and alternately scales it along $x$ and $y$ marginals to produce uniform
marginals in the limit.

**Theorem 1.5.** (a) Given $f \in C(T)$, there exists functions $a_{f,\theta}(\cdot), b_{f,\theta}(\cdot) \in L^1[0,1]$ unique al-
most surely with respect to Lebesgue measure such that $g_{f,\theta}(x,y) := e^{\theta f(x,y) + a_{f,\theta}(x) + b_{f,\theta}(y)}$
is a density with respect to Lebesgue measure on $T$ with uniform marginals. Further,
$\mu_{f,\theta} = g_{f,\theta} dxdy$, and consequently
$$\sup_{\mu \in \mathcal{U}} \{\theta \mu(f) - D(\mu || u)\} = - \int_{x=0}^{1} [a_{f,\theta}(x) + b_{f,\theta}(x)] dx.$$
(b) Consider the maps $S_1, S_2 : \mathcal{C}(T) \to \mathcal{C}(T)$ given by

$$S_1(\phi)(x, y) := \frac{\phi(x, y)}{\int \phi(x, z) dz}, \quad S_2(\phi)(x, y) := \frac{\phi(x, y)}{\int \phi(z, y) dz}.$$ 

Then the sequence starting with $\phi_0 := e^{\theta f(x, y)}$ and given by

$$\phi_{2k+1} = S_1(\phi_{2k}), \quad \phi_{2k+2} = S_2(\phi_{2k+1}), \quad k \geq 0$$

converges in $L^1$ norm to $g_{f, \theta}$.

Remark 1.6. The functions $a_{f, \theta}(\cdot), b_{f, \theta}(\cdot)$ are the unique solutions of the joint integral equations

$$\int_0^1 e^{\theta f(x, z) + a_{f, \theta}(x) + b_{f, \theta}(z)} dz = 1, \quad \int_0^1 e^{\theta f(z, y) + a_{f, \theta}(z) + b_{f, \theta}(y)} dz = 1, \quad \text{for all } x, y \in [0, 1].$$

It follows by part (a) of Theorem 1.4 and Lemma 1.5 that

$$\lim_{n \to \infty} \frac{Z_n(f, \theta) - Z_n(0)}{n} = - \int_{x=0}^1 [a_{f, \theta}(x) + b_{f, \theta}(x)] dx.$$

For the limiting normalizing constant in the Mallows model with the Foot-rule or the Spearman’s coefficient just take $f(x, y) = -|x - y|$ and $f(x, y) = -(x - y)^2$ respectively. Even though analytic computation for $a_{f, \theta}(\cdot), b_{f, \theta}(\cdot)$ might be difficult, the algorithm of Theorem 1.5 can be used for a numerical evaluation of these functions, and hence an estimate of the normalizing constant $Z_n(\theta)$.

Another approach for estimation in such models can be to estimate the parameter $\theta$ without estimating the normalizing constant. The following theorem constructs an explicit $\sqrt{n}$ consistent estimator for $\theta$, for the class of models considered in Theorem 1.4. This estimate is similar in spirit to Besag’s pseudo-likelihood estimator (see [B1],[B2]). In the usual definition the product of conditional distribution of every variable given the rest of the variables is taken as the pseudo-likelihood. Since in a permutation the conditional distribution of $\pi(i)$ given $\{\pi(j), j \neq i\}$ is a point mass, in this case the pseudo-likelihood is taken as

$$\prod_{1 \leq i < j \leq n} Q_{n, f, \theta}(\pi(i), \pi(j)|\pi(k), k \neq i, j),$$

which is the product of conditional distribution of every pair $(\pi(i), \pi(j))$. The equation for estimating $\theta$ constructed in part (b) of Theorem 1.7 can be easily checked to be

$$\sum_{1 \leq i < j \leq n} \frac{\partial}{\partial \theta} \log Q_{n, f, \theta}(\pi(i), \pi(j)|\pi(k), k \neq i, j) = 0.$$

The computation of this estimate requires a grid search over $\mathbb{R}$ and does not require the computation of $Z_n(\theta)$. Thus this gives a fast and practical way for parameter estimation in such models.

Theorem 1.7. (a) In the setting of Theorem 1.4, if

$$f(x_1, y_1) + f(x_2, y_2) - f(x_1, y_2) - f(x_2, y_1) \equiv 0,$$

then $Q_{n, \theta}$ is same as $P_n$ and so there are no consistent estimates for $\theta$. 


Remark 1.8. The condition (1.1) is equivalent to the existence of two functions \( \phi \) and \( \psi \) such that \( f(x, y) = \phi(x) + \psi(y) \). This will be made precise during the proof of Theorem 1.7. In this case \( \sum_{i=1}^{n} f(i/n, \pi(i)/n) \) is a constant free of \( \pi \), and so the probability distribution is same as the uniform distribution \( \mathbb{P}_n \).

As another application of the large deviation principle, in section 4 Theorem 1.9 compute the limiting normalizing constant the from the Mallows model with Kendall’s tau as its metric ((d) in the original list of metrics).

Theorem 1.9. Consider the Mallows model on permutations with Kendall’s tau as the metric, given by

\[
M_n, \theta(\pi) = e^{-\frac{\theta}{2} \sum_{i<j} 1_{\pi(i) > \pi(j)} - Z_n(\theta)},
\]

where \( Z_n(\theta) \) is the appropriate normalizing constant. Then, with \( h : T^2 \to \mathbb{R} \) defined by

\[
h((x_1, y_1), (x_2, y_2)) := 1_{(x_1-x_2)(y_1-y_2)<0} = 1 - 1_{x_1 \geq x_2, y_1 \geq y_2} - 1_{x_1 \leq x_2, y_1 \leq y_2},
\]

(a) \[
\lim_{n \to \infty} \frac{Z_n(\theta) - Z_n(0)}{n} = \sup_{\mu \in \mathcal{U}} \{ -\frac{\theta}{2}(\mu \times \mu)(h) - D(\mu || u) \}.
\]

(b) Under the identification \( \pi \mapsto L_0(\pi) \), \( M_n, \theta \) satisfies a large deviation principle on \( \mathbb{P}(T) \) with rate function \( I_M(\mu) = \inf_{\nu \in \mathcal{U}} I_M(\nu) \), where

\[
I_M(\mu) := D(\mu || u) + \frac{\theta}{2}(\mu \times \mu)(h) \text{ if } \mu \in \mathcal{U},
\]

\( = \infty \text{ otherwise.} \)

Remark 1.10. The fact that the limiting normalizing constant for the Mallows model is given by the above optimization problem was proved in [SS] by a different argument, where the author also derived the weak limit of \( L_0(\pi_n) \) by solving the minimizer of \( I_M \) using a partial differential equation (see the answer in [SS, Theorem 1.1]).

Since in this case the partition function \( Z_n(\theta) \) is explicitly known from [DR], the construction of a consistent estimator is easy. In fact, the maximum likelihood estimate is computable numerically, and it is very easy to check that it is \( \sqrt{n} \) consistent and asymptotically normal.

Section 3.1 carries out some heuristic calculations to approximation the functions \( a_f, \theta(x), b_f, \theta(\cdot) \) when \( \theta \approx 0 \), under the additional assumption that \( f \) is symmetric.

Section 3.2 explores the Spearman’s rank correlation model in some detail, with the location parameter taken to be identity. The density of the limiting measure under this model is plotted...
on a discrete grid using a discretized version of the algorithm proposed in Lemma 1.5. Using a Swendsen-Wang type algorithm from [AD], a random permutation is obtained from this model with \( n = 1000 \). The histogram for the corresponding measure has a very similar pattern as the limiting density plotted above on the grid, thus illustrating the weak convergence result of Theorem 1.4 part (b).

The estimation theory detailed in sections 1-2 relates to estimation of the shape parameter \( \theta \) and take the location parameter as given, not the location parameter \( \pi_0 \). In fact, the estimation of location parameter with one permutation can be a tricky problem. For example, Theorem 1.4 says that under the Spearman’s rank correlation model with identity as the location parameter and \( \theta > 0 \), a random permutation converges to a probability measure on the unit square with positive density everywhere, whereas the location parameter itself converges to the uniform distribution on the diagonal \( x = y \). Thus the two limits are different, and in fact are mutually singular. This does not necessarily imply that estimation of the location parameter cannot be done in this model, but only says that the seemingly reasonable estimator \( \pi_n \) is not a good estimator of \( \pi_0 \), at least in the metric \( d(.,.) \). As a comment, note however that the problem becomes easy when there are i.i.d. replications from this model, and the size of the permutation remains fixed. In this case, the permutation which occurs the most in the sample is a consistent estimate of the true location parameter.

Even though the function \( h \) of Theorem 1.9 is not continuous on \([0,1]^4\), the large deviation result is still applicable, as the functional \( \mu \mapsto (\mu \times \mu)(h) \) is continuous with respect to weak topology. This functional is the natural extension for the number of inversions of a permutation to the space of probability measures on the unit square.

The continuity of this functional follows from [HKMRS, Lemma 5.3], but a separate proof is given in Theorem 1.9. Thus to explore other non uniform models on permutations, one needs to know what are the continuous real valued functionals on \( U \). Any continuous functional on \( U \) can be used as a sufficient statistic for an exponential family on permutations, which can be analyzed using this technique. As an example, note that if \( N(\pi) \) denotes the number of fixed points of \( \pi \), then \( N(\pi)/n \) cannot be extended to a continuous functional on \( U \). Indeed, if \( \pi_n \) denotes the identity permutation on \( S_n \) and \( \sigma_n \) denotes the permutation given by \( \sigma_n(i) = i + 1( \mod n ) \), then both \( \pi_n \) and \( \sigma_n \) converge to the same measure (which is the uniform measure on the diagonal \( x = y \)), but \( N(\pi_n) = n \) and \( N(\sigma_n) = 0 \).

Another interesting problem is to investigate if there is a different topology under which there is a large deviation principle. The topology of weak convergence does not seem to reflect the group structure of \( S_n \). For example even though \( \pi_n \) and \( \sigma_n \) converge to \( \mu \) and \( \nu \) respectively, it might still be true that \( \pi_n \sigma_n \) might not converge. For an explicit example of this take \( \pi_n \) uniform on \( S_n \), and \( \sigma_n = \pi_n \) for \( n \) even and \( \pi_n^{-1} \) for \( n \) odd. In this case \( \pi_n \) and \( \sigma_n \) both converge to \( u \) (Lebesgue measure), but \( \pi_n \sigma_n \) converges to \( u \) along even subsequences, whereas \( \pi_n \sigma_n \) converges to the uniform measure on the diagonal \( x = y \) along odd subsequences. A better topology might take the composition action into account, thus producing more continuous functionals.

1.3. Outline of the paper. In section 2 the Large Deviation Result is stated in Theorem 2.1. This theorem is an easy consequence of [J, Theorem 1]. However, the appendix contains an independent and more direct proof of Theorem 2.1 without using the results from [J], and might be of independent interest. The proof is carried out by using [DZ, Theorem 4.1.11], by choosing a
suitable base for the weak topology.

As an application of Theorem 2.1, section 2 carries out the proofs of Theorems 1.4, 1.5, 1.7 and 1.9. Section 3.1 contains some approximations for \( a_{f, \theta} \) for \( \theta \approx 0 \). Section 3.2 explores the Spearman’s rank correlation model and illustrates the weak convergence of \( \pi_n \) by comparing the theoretical prediction of Theorem 1.4 with a random sample from this distribution.

2. Proofs of main results

The main tool needed for analyzing non uniform models is the following theorem, which gives a Large Deviation Principle under the uniform probability measure on \( S_n \) with respect to the weak topology. The proof of this Theorem has been moved to the appendix.

**Theorem 2.1.** Identifying \( \pi \) with \( L(\pi) \), \( \mathbb{P}_n \) satisfies a large deviation principle on \( \mathcal{U} \) with the good rate function \( I(\mu) := D(\mu||u) \), i.e. for any set \( A \subset \mathcal{U} \),

\[
- \inf_{\mu \in A^o} D(\mu||u) \leq \inf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(A) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(A) \leq - \inf_{\mu \in \overline{A}} D(\mu||u),
\]

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

The following corollary follows immediately from Theorem 2.1. To state the corollary needs a definition.

**Definition 2.2.** For \( \pi \in S_n \), let \( L_0(\pi) \) denote the probability measure \( \sum_{i=1}^{n} \delta_{(i/n, \pi(i)/n)} \), i.e. it puts mass \( 1/n \) at each of the points \((i/n, \pi(i)/n)\). Note that \( L_0(\pi) \) is in \( \mathcal{P}(T) \), but not in \( \mathcal{U} \).

**Corollary 2.3.**

(a) Identifying \( \pi \) with \( L_0(\pi) \), \( \mathbb{P}_n \) satisfies a large deviation principle on \( \mathcal{P}(T) \) with respect to the weak topology, with the good rate function \( \overline{I}(\cdot) \) on \( \mathcal{P}(T) \), given by

\[
\overline{I}(\mu) := D(\mu||u), \quad \mu \in \mathcal{U},
\]

\[
:= \infty \quad \text{otherwise}.
\]

(b) For any \( f \in \mathcal{C}(T) \) (the set of all continuous functions on \( T \)),

\[
\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{P}_n} \left( \sum_{i=1}^{n} f_{(i/n, \pi(i)/n)} \right) = \sup_{\mu \in \mathcal{U}} \{ \mu(f) - D(\mu||u) \}. \tag{2.1}
\]

**Proof.**

(a) Since \( \mathcal{U} \) is closed in \( \mathcal{P}(T) \), [DZ, Lemma 4.1.5 (a)] and Theorem 2.1 gives that under \( \mathbb{P}_n \), the laws of \( L(\pi_n) \) satisfies a large deviation principle on \( \mathcal{P}(T) \) with the rate function \( \overline{I} \).

Now, for any \( f \in \mathcal{T} \) with \( |f|_{\mathcal{L}} \leq 1 \),

\[
|L(\pi_n)(f) - L_0(\pi_n)(f)| \leq \sup_{|x_1 - x_2| \leq \frac{1}{n}, |y_1 - y_2| \leq \frac{1}{n}} |f(x_1, y_2) - f(x_2, y_2)| \leq \frac{\sqrt{2}}{n},
\]

and so for any \( \delta > 0 \), \( d(L(\pi_n), L_0(\pi_n)) \leq \delta \) for all \( n \geq N_\delta \) where \( N_\delta \) is non-random. Thus by [DZ, Theorem 4.2.13], \( L_0(\pi_n) \) also satisfies the same large deviation principle as \( L(\pi_n) \), completing the proof of part (a).

(b) The map \( \mu \mapsto \mu(f) \) is bounded and continuous with respect to weak topology, and so an application of Varadhan’s lemma ([DZ, Theorem 4.3.1]) gives (2.1). □
Proof of Theorem 1.4. (a) Note that
\[ e^{Z_n(f,\theta) - Z_n(0)} = \frac{1}{n!} \sum_{\pi \in S_n} e^{\theta \sum_{i=1}^n f(i/n, \pi_i/n)} = \mathbb{E}_{\pi_0} e^{\theta \sum_{i=1}^n f(i/n, \pi_i/n)}, \]
since \( Z_n(0) = \log n! \). Part (a) follows from applying (2.1) with \( f(.) \) replaced by \( \theta f(.) \).

(b) It follows by [Ho, Theorem III.17] and part (a) of Corollary 2.3 that \( L_0(\pi_n) \) satisfies a large deviation principle with the good rate function \( I_{f,\theta}(\cdot) \) given by
\[ I_{f,\theta}(\mu) := D(\mu||u) - \theta \mu(f) - \inf_{\mu \in S} \{D(\mu||u) - \theta \mu(f)\}, \text{ if } \mu \in S, \]
\[ = \infty \text{ otherwise.} \]

Since \( I_{f,\theta} \) is strictly convex and \( U \) is closed, there exists a unique minimizer \( \mu_{f,\theta} \) of \( I_{f,\theta}(\cdot) \) in \( U \).

Now the set \( \{ \mu : d_{L}(\mu, \mu_{f,\theta}) \geq \epsilon \} \) is closed, and so the large deviation for \( L_0(\pi_n) \) gives
\[ \lim_{n \to \infty} \frac{1}{n} \log \mathbb{Q}_{n,f,\theta}(d_{L}(L_0(\pi_n), \mu_{f,\theta}) \geq \epsilon) \leq - \inf_{\mu \in S ; d_{L}(\mu, \mu_{f,\theta}) \geq \epsilon} I_{f,\theta}(\mu) + \inf_{\mu \in S} I_{f,\theta}(\mu). \]
The quantity on the r.h.s. above is negative as \( I_{f,\theta} \) is a lower semi continuous function, and \( \{ \mu : d_{L}(\mu, \mu_{f,\theta}) \geq \epsilon \} \) is compact. This proves the weak convergence of \( L_0(\pi_n) \).

Proof of Lemma 1.5. (a) Since \( \theta f(.) \) is integrable with respect to \( du \), by [Cs, Corollary 3.2] there exists functions \( a(.) , b(.) : L^1[0,1] \) such that
\[ d_{\mu_{a,b}} = g_{a,b}dxdy := e^{\theta f(x,y) + a(x) + b(y)} dxdy \in U. \]
Note that the dependence of \( a(.) , b(.) \) on \( f, \theta \) is not made explicit for convenience, as both \( \theta, f \) remain fixed throughout the argument. The proof of \( \mu_{a,b} = \mu_{f,\theta} \) is by way of contradiction. Suppose this is not true. Since \( \mu_{f,\theta} \) is the unique global minimizer of \( I_{f,\theta} \), 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
\[ I_{f,\theta}(\mu_{f,\theta}) \leq I_{f,\theta}(u) = -\theta u(f) < \infty, \]
which forces \( D(\mu_{f,\theta}||u) < \infty \). Thus letting \( \phi_{f,\theta} := \frac{d\mu_{f,\theta}}{du} \) gives
\[ h'(0) = \int_T (\phi_{f,\theta}(x,y) - g_{a,b}(x,y))(\log g_{a,b}(x,y) - \theta f(x,y))du \]
\[ = \int_T (\phi_{f,\theta}(x,y) - g_{a,b}(x,y))(a(x) + b(y))du \]
\[ = \mathbb{E}_{\mu_{f,\theta}}[a(X) + b(Y)] - \mathbb{E}_{\mu_{a,b}}[a(X) + b(Y)] = 0, \]
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 \( \arg \min_{\mu \in U} I_{f,\theta}(\mu) \) which contradicts part (b) of Theorem 1.4. Thus \( d\mu_{f,\theta} = d\mu_{a,b} = e^{\theta f(x,y) + a(x) + b(y)} dxdy \). Finally, the almost sure uniqueness of \( a(.) \) and \( b(.) \) follows from the uniqueness of the measure \( \mu_{f,\theta} \).
Using this and the definition of \( I_{f,\theta} \) the last claim of part (a) follows trivially.
(b) Since \( \phi_0 \) is continuous and strictly positive, it follows by induction that \( \phi_k \) is well defined and strictly positive, and is of the form \( e^{f(x,y) + a_k(x) + b_k(y)} \) with \( a_k, b_k \in C[0,1] \). The convergence in \( L^1 \) then follows from ([Kb]).

\[ \]

Before proving Theorem 1.7, a general lemma is stated which constructs \( \sqrt{n} \) consistent estimates of \( \theta \) in one parameter families.

**Lemma 2.4.** Let \( \mathbb{R}_{n,\theta} \) be a one parameter family over any \( \sigma \) finite measure space \((\mathcal{X}_n, \mathcal{F}_n)\) with \( \theta \in \mathbb{R} \), and let \( \theta_0 \in \mathbb{R} \) be fixed. Let \( P_n(x, \theta) \) be a function which is measurable in \( x \) for fixed \( \theta \), and differentiable in \( \theta \), such that the following two conditions hold:

(a) There exists \( C = C(\theta_0) < \infty \) such that

\[
\mathbb{E}_{\mathbb{R}_{n,\theta_0}} P_n(x, \theta_0)^2 \leq \frac{C}{n} \tag{2.2}
\]

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

\[
\lim_{n \to \infty} \mathbb{R}_{n,\theta_0}(P_n(x, \theta) \leq -\lambda(\theta), \theta \in \mathbb{R}) = 1 \tag{2.3}
\]

Then the equation \( P_n(x, \theta) = 0 \) has a unique root in \( \theta \). Further, this root by \( \hat{\theta}_n \) has the property that \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) is \( O_p(1) \) under \( \mathbb{R}_{n,\theta_0} \).

**Proof.** For any sequence \( M_n \to \infty \) such that \( M_n = o(\sqrt{n}) \),

\[
\mathbb{R}_{n,\theta_0}(|P_n(x, \theta_0)| > M_n/\sqrt{n}) \leq \frac{n}{M_n^2} \mathbb{E}_{\mathbb{R}_{n,\theta_0}} P_n(x, \theta_0)^2 \leq \frac{C}{M_n^2} \to 0.
\]

This along with (2.3) gives \( Q_{n,\theta_0}(A_n) \to 1 \), where

\[
A_n := \{x \in \mathcal{X}_n : |P_n(x, \theta_0)| \leq M_n/\sqrt{n}, \ P_n'(x, \theta) \leq -\lambda(\theta), \ \theta \in \mathbb{R}\}.
\]

For \( x \in A_n \) and any \( \delta > 0 \),

\[
P_n(x)(\theta_0 + \delta) = P_n(x, \theta_0) + \int_{\theta = \theta_0}^{\theta_0 + \delta} P_n'(x, \theta)d\theta \leq \frac{M_n}{\sqrt{n}} - \delta \inf_{\theta \in [\theta_0, \theta_0 + \delta]} \lambda(\theta) < 0
\]

for all large \( n \). Similarly it can be shown that \( P_n(x, \theta_0 - \delta) > 0 \) for all large \( n \). Also note that since \( x \in A_n \), \( P_n(x, \theta) \) is monotone, and so by continuity of \( P_n(x, \theta) \) there exists unique \( \hat{\theta}_n \) satisfying \( P_n(x, \hat{\theta}_n) = 0 \), and \( \theta_0 - \delta < \hat{\theta}_n < \theta_0 + \delta \) for all large \( n \). Thus,

\[
\frac{M_n}{\sqrt{n}} \geq |P_n(x, \theta_0)| = |P_n(x, \theta_0) - P_n(x, \hat{\theta}_n)| = \int_{\theta = \theta_0}^{\theta_0 + \delta} P_n'(x, \theta)d\theta \geq \int_{\theta = \theta_0}^{\theta_0 + \delta} \lambda(\theta)d\theta \geq \inf_{[\theta-\theta_0] \leq \delta} \lambda(\theta) |\hat{\theta}_n - \theta_0|,
\]

and so \( \mathbb{R}_{n,\theta_0}(\sqrt{n}|\hat{\theta}_n - \theta_0| \leq K(\theta_0)M_n) \to 1 \) for \( K(\theta_0) := \inf_{[\theta-\theta_0] \leq \delta} \lambda(\theta) \) \( -1 < \infty \), for any \( M_n \to \infty \), concluding the proof of the Lemma.

**Proof of Theorem 1.7.** (a) If (1.1) holds, then fixing \( y \in [0,1] \) and setting \( H(x, y) := f(x, y) - f(x, 0) \), for any \( x_1, x_2 \in [0,1] \),

\[
H(x_1, y) - H(x_2, y) = f(x_1, y) - f(x_1, 0) - f(x_2, y) + f(x_2, 0) = 0,
\]

and so \( H(x, y) \) is a constant function in \( x \). Thus denoting \( H(x, y) \) by \( \psi(y) \) \( f(x, y) \) can be written as

\[
f(x, y) = f(x, 0) + \psi(y) = \phi(x) + \psi(y), \quad \phi(x) := f(x, 0).
\]
and so
\[ \sum_{i=1}^{n} f(i/n, \pi(i)/n) = \sum_{i=1}^{n} \phi(i/n) + \sum_{i=1}^{n} \psi(i/n) = x_n \]
for some deterministic sequence of reals \( x_n \) free of \( \pi \). The the distribution \( Q_{n,\theta} \) is same as \( P_n \), and consequently no consistent estimator for \( \theta \) exists.

(b) It suffices to check the two conditions of Lemma 2.4 with \( \mathbb{R}_{n,\theta} = Q_{n,\theta}, \mathcal{X}_n = S_n \). For checking (2.2) an exchangeable pair is constructed. The idea of the proof is taken from [Ch, Lemma 1.2].

Consider the following exchangeable pair of permutations \((\pi, \pi')\) on \( S_n \) constructed as follows: Pick \( \pi \) from \( Q_{n,\theta} \). For constructing \( \pi' \), first pick \((I < J)\) uniformly from \( \{(i, j) : 1 \leq i < j \leq n\} \), and let
\[
(\pi'(I), \pi'(J)) = (\pi(I), \pi(J)) \quad \text{w.p.} \quad Q_{n,\theta}(\pi(I') = \pi(I), \pi(J') = \pi(J)|\pi(k), k \neq I, J)
\]
\[= e^{\theta a(I, I, \pi) + \theta a(J, J, \pi)} \cdot \left( e^{\theta a(I, J, \pi) + \theta a(J, I, \pi)}, e^{\theta a(J, I, \pi) + \theta a(J, J, \pi)} \right) + e^{\theta a(I, J, \pi) + \theta a(J, I, \pi)} \cdot \left( e^{\theta a(I, I, \pi) + \theta a(J, J, \pi)} \right),
\]
and set \( \pi'(i) = \pi(i) \) for all \( i \neq I, J \). It can be readily checked that \((\pi, \pi')\) is indeed an exchangeable pair. Thus defining
\[
W(\pi) := \sum_{i=1}^{n} f(i/n, \pi(i)/n), \quad F(\pi, \pi') := W(\pi) - W(\pi')
\]
it follows from the construction of \((\pi, \pi')\) that \( E_{Q_{n,\theta}}[F(\pi, \pi')|\pi] = P_n(\pi, \theta) \). Also,
\[
E_{Q_{n,\theta}} P_n(\pi, \theta)^2 = E_{Q_{n,\theta}} P_n(\pi, \theta) F(\pi, \pi') = E_{Q_{n,\theta}} P_n(\pi, \theta) F(\pi', \pi) = -E_{Q_{n,\theta}} P_n(\pi, \theta) F(\pi, \pi'),
\]
where it has been used that \((\pi, \pi')\) are exchangeable and \( F \) is antisymmetric. This readily implies
\[
E_{Q_{n,\theta}} P_n(\pi, \theta)^2 = \frac{1}{2} E_{Q_{n,\theta}} E_{Q_{n,\theta}}[(P_n(\pi, \theta) - P_n(\pi', \theta))F(\pi, \pi')|\pi] =: E_{Q_{n,\theta}} V_n(\pi). \quad (2.4)
\]
Letting \( \pi^{ij} \) denote \( \pi \) with the elements \((\pi(i), \pi(j))\) swapped, \( V_n(\pi) \) can be written as
\[
\frac{1}{2N_n} \sum_{1 \leq i < j \leq n} \left[ P_n(\pi, \theta) - P_n(\pi^{ij}, \theta) \right] \left[ y(i, i, \pi) + y(j, j, \pi) - y(i, j, \pi) - y(j, i, \pi) \right] Q_{n,\theta}((\pi'(i), \pi'(j)) = (\pi(j), \pi(i))|\pi),
\]
Also note that
\[
|P_n(\pi, \theta) - P_n(\pi^{ij}, \theta)| \leq \frac{2nM}{N_n}, \quad \text{where} \quad M := 4 \sup_T |f|,
\]
and so \( |V_n(\pi)| \leq \frac{nM^2}{N_n} \), which, along with (2.4), completes the proof of (2.2) with \( C = 3M^2 \).
Proceeding to check (2.3), note that setting $c_{ij} := y(i,i,\pi) + y(j,j,\pi)$ and $d_{ij} := y(i,j,\pi) + y(j,i,\pi)$ gives that

$$-P_n(\pi,\theta)^2 = \frac{1}{N_n^2} \sum_{1 \leq i < j \leq n} (c_{ij} - d_{ij})^2 \frac{e^{\theta(c_{ij} + d_{ij})}}{(e^{\theta c_{ij}} + e^{\theta d_{ij}})^2} \geq \frac{e^{-|\theta|M}}{8N_n^2} \sum_{i,j=1}^n (c_{ij} - d_{ij})^2,$$

where the last inequality uses the fact that $|c_{ij}| \leq M/2, |d_{ij}| \leq M/2$ by definition of $M$. Since the function $g : T \times T \mapsto \mathbb{R}$ defined by

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

is continuous on $T^2$,

$$\frac{1}{2} \sum_{i=1}^n (c_{ij} - d_{ij})^2 = \frac{1}{n^2} \sum_{i,j=1}^n g((i/n, \pi(i)/n), (j/n, \pi(j)/n)) = \left( L_0(\pi_n) \times L_0(\pi_n) \right)(g) \xrightarrow{P} \int_{T^2} \left[ f(x_1, y_1) + f(x_2, y_2) - f(x_1, y_2) - f(x_2, y_1) \right]^2 d\mu_{f,\theta}(x_1, y_1) d\mu_{f,\theta}(x_2, y_2).$$

To check the last convergence, note that $L_0(\pi_n) \Rightarrow \mu_{f,\theta}$ in probability by part (b) of Theorem 1.4, which readily gives $L_0(\pi_n) \times L_0(\pi_n) \Rightarrow \mu_{f,\theta} \times \mu_{f,\theta}$ in probability as well. Finally note that the last integral is a strictly positive real $\alpha > 0$ by (1.1), and so (2.3) holds with $h(\theta) = e^{-M|\theta|\alpha/5}$, and so by Lemma 2.4 the conclusion follows.

\[ \square \]

**Proof of Theorem 1.9.** (a) First it will be shown that $\mu \mapsto -\theta(\mu \times \mu)(h)/2$ is continuous with respect to weak topology on $\mathcal{U}$. Since $\mathcal{U}$ is separable, it suffices to work with sequences, and it suffices to check the following:

$$\mu_k \in \mathcal{U}, \mu_k \xrightarrow{w} \mu \Rightarrow (\mu_k \times \mu_k)(x_1 \leq x_2, y_1 \leq y_2) \rightarrow \mu(x_1 \leq x_2, y_1 \leq y_2)$$

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

Now, a similar computation as in the proof of Theorem 1.4 gives

$$e^{Z_n(\theta) - Z_n(0)} = \frac{1}{n!} \sum_{\pi \in S_n} e^{-\frac{\alpha}{n}} \sum_{1 \leq i < j \leq n} h((i/n, \pi(i)/n), (j/n, \pi(j)/n)) = \mathbb{E}_{\pi_n} e^{-\frac{\alpha}{2} (L_0(\pi) \times L_0(\pi))(h)},$$

where

$$0 \leq h((x_1, y_1), (x_2, y_2)) = 1_{(x_1-x_2)(y_1-y_2) > 1} \leq 1.$$

Part (a) then follows by an application of Varadhan's Lemma along with Corollary 2.3, 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 $\{\mathcal{T}(\mu) < \infty\}$, which is true. (See also part (c) of the Remark following [DZ, Theorem 4.2.1].)
b) Let $\psi$ be any bounded continuous linear functional on $\mathbb{P}(T)$. It will be shown that
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{P}_n, \theta} \mathbb{E}^{n\psi(L_0(\pi))} = \sup_{\mu \in \mathcal{U}} \{\psi(\mu) - \theta \mathbb{E}^{(\mu \times \mu)(h)} - D(\mu||u)\} - \sup_{\mu \in \mathcal{U}} \{\frac{\theta}{2} (\mu \times \mu)(h) - D(\mu||u)\}
\]
Indeed, using part (a), this is equivalent to showing
\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{P}_n} \mathbb{E}^{n\psi(L_0(\pi_n)) - n\frac{2}{\theta} L_0(\pi_n) \times L_0(\pi_n)(h)} = \sup_{\mu \in \mathcal{U}} \{\psi(\mu) - \theta \mathbb{E}^{(\mu \times \mu)(h)} - D(\mu||u)\},
\]
which follows by another application of Varadhan’s lemma.

Finally, since the space $\mathbb{P}(T)$ is compact, an application of Bryc’s inverse Varadhan Lemma ([DZ, Theorem 4.4.2]) gives that $\mathbb{M}_n, \theta$ satisfies a large deviation principle with rate function $I_{\mathbb{M}}(\mu) = \inf_{\nu \in \mathcal{U}} I_{\mathbb{M}}(\nu)$ for $\mu \in \mathcal{U}$, and $I_{\mathbb{M}}(\mu) = \infty$ for $\mu \notin \mathcal{U}$.

3. Some approximations and an example

3.1. Approximations for small $\theta$. This sub-section uses heuristic methods to explore the model of Theorem 1.4 for $\theta \approx 0$.

Let $f$ be a symmetric continuous function on the unit square, and consider the model $\mathbb{Q}_{n,f,\theta}$ as introduced in Theorem 1.4. Then by Lemma 1.5, a randomly chosen $\pi_n$ from this model converges weakly to a measure on $T$ with density of the form $e^{\theta f(x,y)}h_{f,\theta}(x)h_{f,\theta}(y)$ with respect to Lebesgue measure, where $h_{f,\theta}(x) := e^{a_{f,\theta}(x)}$. Also note that $h_{f,0}(x) \equiv 1$ which gives $a_{f,0}(x) \equiv 0$. Thus assuming that $h_{f,\theta}$ is smooth in $\theta$, using one term Taylor approximation $h_{f,\theta}(x) = 1 + \theta h'_{f,\theta}(x) + O(\theta^2)$ and the uniform marginal condition gives

\[
1 = 1 + \theta h'_{f,0}(x) + \theta \int_{y=0}^{1} f(x,y)dy + \theta \int_{y=0}^{1} h'_{f,0}(y)dy + O(\theta^2)
\]

\[
\Rightarrow h'_{f,0}(x) = -\int_{y=0}^{1} f(x,y)dy - C,
\]
where $C = \int_{y=0}^{1} h'_{f,0}(y)dy$. Integrating the above equation gives $2C = -\int_{x,y=0}^{1} f(x,y)dy$. This gives the following approximation for $a_{f,\theta}(x)$ and the log partition function:

\[
a_{f,\theta}(x) = -\theta \int_{y=0}^{1} f(x,y)dy + \frac{\theta}{2} \int_{x,y=0}^{1} f(x,y)dxdy + O(\theta^2)
\]

\[
\frac{1}{n} [Z_n(f, \theta) - Z_n(0)] = \theta \int_{x,y=0}^{1} f(x,y)dxdy + O(\theta^2) + o_n(1)
\]
The second approximation also follows from exponential family results, as by a one term Taylor’s expansion of $Z_n(f, \theta) - Z_n(0) = \theta Z'_n(f, 0) + O(\theta^2)$ with

$$\frac{Z'_n(f, 0)}{n} = \frac{1}{n} \sum_{i=1}^{n} f\left(\frac{i}{n}, \frac{\pi_i}{n}\right) p_{\pi_i} \int f(x, y) dx dy$$

under uniform measure on $S_n$.

3.2. One particular model: The Spearman’s rank correlation. The Spearman’s rank correlation model is obtained by setting $f(x, y) = -(x - y)^2$ in the model of Theorem 1.4, and is given by

$$Q_{n, \theta}(\pi) = e^{(\theta/n^2) \sum_{i=1}^{n} (i - \pi(i))^2 - Z_n(\theta)}.$$

This can also be written as

$$e^{(\theta/n^2) \sum_{i=1}^{n} i \pi(i) - W_n(\theta)}$$

for some different normalizing constant $W_n(\theta)$, and so Theorem 1.4 works with $f(x, y) = xy$ as well. This subsection will work with the choice $f(x, y) = xy$ for convenience. Since $f$ is fixed, the dependence of $a_{\theta}(.)$, $h_{\theta}(.)$ and $Q_{n, \theta}$ on $f$ is not made explicit in this section.

Setting $h_{\theta}(x) := e^{a_{\theta}(x)}$ as in section 3.1 and using the uniform marginal condition gives

$$1 = \int_{y=0}^{1} e^{\theta xy} h_{\theta}(x) h_{\theta}(y) = h_{\theta}(x) \sum_{k=0}^{\infty} C_k(\theta) \frac{x^k \theta^k}{k!}$$

with $C_k(\theta) := \int_{y=0}^{1} y^k h_{\theta}(y) dy$, and so

$$h_{\theta}(x) = \left(\sum_{k=0}^{\infty} C_k(\theta) \frac{x^k \theta^k}{k!}\right)^{-1}.$$

Another integration with respect to $x$ gives

$$\sum_{k=0}^{\infty} \frac{C_k^2(\theta)}{k!} = 1.$$

However, analytic solution of $a_{\theta}(.)$ seems intractable and is not attempted here.

Instead, below is plotted the limiting function $g_{\theta}(x, y) := e^{a_{\theta}(x)+a_{\theta}(y)}$ obtained from this model, on a discrete grid of size $k \times k$ with $k = 1000$. The function is computed by iterative scaling of row and column sums of a $k \times k$ matrix $A$ starting with $A(i, j) = e^{(\theta/n^2)ij}$, where $\theta = 20$. This is the discrete version of the algorithm of Lemma 1.5.
From the figure it is easy to see that $g_\theta$ has higher values on the diagonal $x = y$, which also follows from the fact that for $\theta > 0$ the identity permutation has the largest probability under this model. The function $g_\theta(.,.)$ is symmetric about the diagonal $x = y$, which is obvious since $f(.,.)$ is symmetric. Another way to see this is by noting that if $\pi_n$ converges to a probability measure on $T$ with limiting density $g_\theta(x,y)$, then $\pi_n^{-1}$ converges to a measure with limiting density $g_\theta(y,x)$. But since
\[
\sum_{i=1}^{n} i\pi(i) = \sum_{i=1}^{n} i\pi^{-1}(i),
\]
the law of $\pi_n$ and $\pi_n^{-1}$ are same under $Q_{n,\theta}$, and so $\pi_n^{-1}$ has the limiting density $g_\theta(x,y)$ as well, thus giving $g_\theta(x,y) = g_\theta(y,x)$.

The function is also symmetric about the other diagonal $x + y = 1$. A similar reasoning as above justifies this:

Define $\sigma \in S_n$ by $\sigma_i := n + 1 - \pi^{-1}(n + 1 - \pi_i)$ and note that if $\pi$ converges to a probability on $T$ with density $g_\theta(x,y)$, then $\sigma$ converges to a probability on $T$ with density $g_\theta(1 - y, 1 - x)$. But
since
\[ \sum_{i=1}^{n} i \pi_i = \sum_{i=1}^{n} (n + 1 - i)(n + 1 - \pi_i) = \sum_{i=1}^{n} i \sigma_i, \]
it follows that under $Q_{n, \theta}$ the distribution of $\pi$ is same as the distribution of $\sigma$. Thus $\sigma$ has limiting density $g_{\theta}(x, y)$, which implies $g_{\theta}(x, y) = g_{\theta}(1 - y, 1 - x)$, and so $g_{\theta}$ is symmetric about the line $x + y = 1$.

To compare how close a random permutation from this model is to the limit, a permutation of size $n = 10000$ is drawn from this model via MCMC. The algorithm used to simulate from this model is taken from [AD], and is explained below:

1. Start with $\pi$ chosen uniformly at random from $S_n$.
2. Given $\pi$, simulate $\{U_i\}_{i=1}^{n}$ mutually independent with $U_i$ uniform on $[0, e^{(\theta/n^2)i\pi_i}]$.
3. Given $U$, let $b_j := n^2 \log U_j / \theta$. Then $1 \leq b_j \leq n$. Choose an index $i_1$ uniformly at random from set $\{j : b_j \leq 1\}$, and set $\pi_{i_1} = 1$. Remove this index from $[n]$ and choose an index $i_2$ uniformly from $\{j : b_j \leq 2\} - \{i_1\}$, and set $\sigma_{i_2} = 2$. In general, having defined $\{i_1, \cdots, i_{l-1}\}$, remove them from $[n]$, and choose $i_l$ uniformly from $\{j : b_j \leq l\} - \{i_1, i_2, \cdots, i_{l-1}\}$, and set $\pi_{i_l} = l$. [That this can be always done was proved in [DGH].]
4. Iterate between the steps 2 and 3 till convergence.

The above iteration is run 10 times to obtain a single permutation $\pi$, and then the frequency histogram of the points $\{t/n, \pi(t)/n\}_{t=1}^{n}$ are computed with $k \times k$ bins, where $k = 10$. The mesh plot of the above frequency histogram is given below.
The pattern of the histogram in Figure 1 is very similar to the function plotted in Figure 1, showing that the probability assigned by the random permutation $\pi$ from this model to squares of size $0.1 \times 0.1$ (since $k = 10$) have a similar pattern as that of the limiting density $g_\theta(x, y)$. 

**Figure 2.** Mesh plot for random permutation from Spearman’s model with $n = 10000$
4. Acknowledgements

I would like to thank my advisor Dr. Persi Diaconis for his inspirational guidance during my Ph.D, and Dr. Amir Dembo for helpful discussions on this paper. I would also like to thank Dr. Sourav Chatterjee, Dr Susan Holmes, Dr. S.R.S. Varadhan, Bhaswar Bhattacharya and Austen Head for suggestions and comments.

References

[AD] H. Andersen and P. Diaconis, Hit and run as a unifying device. Journal de la Societe Francaise de Statistique, 148(4):5-28, (2007).

[B1] J. Besag, Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological). 36 (2): 192-236, (1974).

[B2] J. Besag, Statistical Analysis of Non-Lattice Data. Journal of the Royal Statistical Society. Series D (The Statistician). 24 (3): 179-195 (1975)

[BCLSSV] C. Borgs, J. Chayes, L. Lovasz, V. T. Sos, B. Szegedy, and K. Vesztergombi, Graph Limits and Parameter Testing. Proceedings of the Annual ACM Symposium on Theory of Computing, 38: 261-270. (2006)

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

[Cs] I. Csiszar, I-Divergence Geometry of Probability Distributions and Minimization Problems. The Annals of Probability. 3 (1): 146-158. (1975)

[D] P. Diaconis, Group representations in probability and statistics. Institute of Mathematical Statistics, Lecture Notes-Monograph series 11. (1988)

[DGH] P. Diaconis, R. Graham, and S. Holmes, Statistical Problems Involving Permutations with Restricted Positions, Lecture Notes-Monograph Series. 36: 195-222. (2001)

[DR] P. Diaconis and A. Ram, Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. Michigan Journal of Mathematics, 48(1):157-190. (2000)

[Dudley] R. Dudley, Real analysis and probability. Cambridge University Press.(2002)

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

[GGKK] R. Glebov, A. Grzesik, T. Klimosova, and D. Kral, Finitely forcible graphons and permutations. arXiv:1307.2444. (2013)

[HKMRS] C. Hoppen, Y. Kohayakawa, C.G.Moreira, B.Rath, and I.M. Sampaio, Limits of permutation sequences. Journal of Combinatorial Theory Series B. 96 (6): 933-957. (2009)

[Ho] F. den. Hollander, Large deviations. Fields Institute Monographs, American Mathematical Society, Providence, RI. (2000)

[J] J. Trashorras, Large Deviations for Symmetrised Empirical Measures. Journal of Theoretical Probability. 21, 397-412. (2008)

[Kb] S. Kullback, Probability Densities with Given Marginals. The Annals of Mathematical Statistics. 39 (4): 1236-1243. (1968)

[KK] S. Kullback and M. A. Khairat, A Note on Minimum Discrimination Information. The Annals of Mathematical Statistics. 37 (1): 279-280. 1966.

[LS] L. Lovasz and B. Szegedy, Limits of dense graph sequences. Journal of Combinatorial Theory. Series B. 96 (6): 933-957. (2006)

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

5. Appendix: Proof of Theorem 2.1

At first a few auxiliary lemmas will be proved, and then Theorem 2.1 will be derived as a consequence of these lemmas. The following two definitions are needed for stating the first lemma.

Definition 5.1. For $k \in \mathbb{N}$, partition $T$ into $k^2$ unit squares $\{T_{rs}\}_{r,s=1}^{k}$, with

$$T_{rs} := \{(x,y) \in T : \frac{r-1}{k} < x \leq \frac{r}{k}, \frac{s-1}{k} < y \leq \frac{s}{k}\}.$$
By definition $T = \bigcup_{s=1}^{k} T_{rs} \cup \{(0, y) : 0 \leq y \leq 1\} \cup \{(x, 0) : 0 \leq x \leq 1\}$, and the last two sets have 0 probability under any $\mu \in \mathcal{U}$. Let $(a_r, b_s) := \frac{1}{k} (r - .5, s - .5)$ denote the center of $T_{rs}$, and define the map $P_k : \mathcal{U} \mapsto \mathbb{P}(T)$ by setting $P_k(\mu)$ to be the probability measure on $T$ which puts mass $\mu(T_{rs})$ at $(a_r, b_s)$, for $1 \leq r, s \leq k$.

**Definition 5.2.** Let

$$A_{k,n} := \left\{ \{m_{rs}\}_{r,s=1}^{k} \in (\mathbb{N} \cup \{0\})^{k} : \sum_{s=1}^{k} m_{rs} = \left\lfloor \frac{nr}{k} \right\rfloor - \left\lfloor \frac{n(r-1)}{k} \right\rfloor, \sum_{r=1}^{k} m_{rs} = \left\lfloor \frac{ns}{k} \right\rfloor - \left\lfloor \frac{n(s-1)}{k} \right\rfloor \right\}.$$ 

In words, $A_{k,n}$ denotes the number of non negative integer valued $k \times k$ matrices with given row sums $\left\lfloor \frac{nr}{k} \right\rfloor - \left\lfloor \frac{n(r-1)}{k} \right\rfloor$ and given column sums $\left\lfloor \frac{ns}{k} \right\rfloor - \left\lfloor \frac{n(s-1)}{k} \right\rfloor$. Note that any $\{m_{rs}\}_{r,s=1}^{k} \in A_{k,n}$ satisfies $\sum_{r,s=1}^{k} m_{rs} = n$.

**Lemma 5.3.** For $1 \leq r, s \leq k$ let $T_{rs}$ be as in (5.1), and let $M_{rs} = M_{rs}(\pi) := \{i \in [n] : (i/n, \pi_i/n) \in T_{rs}\}$.

Then the joint distribution of $\{M_{r,s}\}_{r,s=1}^{k}$ is given by

$$\mathbb{P}_n(M_{r,s} = m_{r,s}, 1 \leq r, s \leq k) = \frac{\prod_{r=1}^{k} m_{r}! \prod_{s=1}^{k} m_{s}!}{n! \prod_{r,s=1}^{k} m_{rs}!}, \quad m_r = \sum_{s=1}^{k} m_{rs}, m_s = \sum_{r=1}^{k} m_{rs}$$

if $m_r = \left\lfloor \frac{nr}{k} \right\rfloor - \left\lfloor \frac{n(r-1)}{k} \right\rfloor, m_s = \left\lfloor \frac{ns}{k} \right\rfloor - \left\lfloor \frac{n(s-1)}{k} \right\rfloor$, and 0 otherwise.

**Proof.** Since

$$M_{r,s} = \#\{i : \frac{r-1}{k} < \frac{i}{n} \leq \frac{r}{k}, \frac{s-1}{k} < \frac{\pi_i}{n} \leq \frac{s}{k}\},$$

it follows that

$$\sum_{s=1}^{k} M_{r,s} = \#\{i : \frac{r-1}{k} < \frac{i}{n} \leq \frac{r}{k}\}, \quad \sum_{r=1}^{k} M_{r,s} = \#\{i : \frac{s-1}{k} < \frac{\pi_i}{n} \leq \frac{s}{k}\},$$

and so any valid configuration $\{M_{r,s}\}_{r,s=1}^{k}$ must be in $A_{k,n}$. So fix $\{m_{rs}\}_{r,s=1}^{k} \in A_{k,n}$. The number of possible permutations with these specified values of $\{m_{rs}\}_{r,s=1}^{k}$ 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 $\{T_{rs}\}_{r,s=1}^{k}$ in $m_r! \prod_{s=1}^{k} m_{rs}!$ ways, so that box $T_{rs}$ 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 the \( T_{r,s} \) receives \( m_{rs} \) targets is

\[
\prod_{s=1}^{k} m_s! \prod_{r,s=1}^{k} m_{rs}!
\]

Finally, after the above distribution, each box \( T_{r,s} \) has \( m_{rs} \) indices and \( m_{rs} \) targets, which can be permuted freely, and so the total number of permutations coming out of one such distribution of indices and targets is

\[
\prod_{r,s=1}^{k} m_{rs}!
\]

Combining, the total number of possible permutations from a given specification of \( \{ m_{rs} \} \) is given by

\[
\prod_{r=1}^{k} m_r! \prod_{s=1}^{k} m_s! \prod_{r,s=1}^{k} m_{rs}!
\]

Since the total number of permutations in \( n! \), the proof of the claim is complete. \( \square \)

Before proceeding the following definitions are needed. The first definition gives a base for the weak topology on \( U \). The second definition introduces a few notations needed for stating the next two lemmas, which are the main steps for proving the LDP for \( \mathbb{P}_n \).

**Definition 5.4.** For \( k, m \in \mathbb{N}, \epsilon > 0, f_m := (f_1, f_2, \cdots, f_m) \) with \( f_i \in C(T) \) and \( \mu_0 \in U \), define

\[
U[k, m, \epsilon, f_m, \mu_0] := \{ \mu \in U : |P_k(\mu)(f) - P_k(\mu_0)(f)| < \epsilon \}
\]

and define

\[
U_0 := \{ U[k, m, \epsilon, f_m, \mu_0] : k, m \in \mathbb{N}; \epsilon > 0; \{ f_i \}_{i=1}^{m} \subset C(T); \mu_0 \in T \}
\]

It is easy to check that \( U_0 \) is a base for the weak convergence on \( U \).

**Definition 5.5.** Let

\[
A_k := \{ \{ x_{rs} \}_{r,s=1}^{k} : x_{rs} \geq 0, \sum_{r=1}^{k} x_{rs} = 1/k, \sum_{s=1}^{k} x_{rs} = 1/k \} \subset [0, 1]^{k^2},
\]

\[
V_n[k, m, f_m, \mu_0](\epsilon) := \{ \{ m_{rs} \}_{r,s=1}^{k} \in A_{k,n} : \left| \sum_{r,s=1}^{k} \left( \frac{m_{rs}}{n} - P_k(\mu_0)(a_r, b_s) \right) f_i(a_r, b_s) \right| < \epsilon, 1 \leq i \leq m \} \subset (\mathbb{N} \cup \{0\})^{k^2},
\]

\[
V[k, m, f_m, \mu_0] := \{ \{ x_{rs} \}_{r,s=1}^{k} \in A_k : \left| \sum_{r,s=1}^{k} (x_{rs} - P_k(\mu_0)(a_r, b_s)) f_i(a_r, b_s) \right| < \epsilon, 1 \leq i \leq m \} \subset [0, 1]^{k^2}.
\]

**Lemma 5.6.**

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n (\{ M_{rs} \}_{r,s=1}^{k} \in V_n[k, m, f_m, \mu_0](\epsilon)) = -2 \log k - \inf_{V[k, m, f_m, \mu_0](\epsilon)} H(\{ x_{rs} \}),
\]

where \( H(\{ x_{rs} \}) := \sum_{r,s=1}^{k} x_{rs} \log x_{rs} \).
Proof. For the proof, first assume that

\[ \lim_{n \to \infty} \min_{\mathcal{V}_n[k,m,f_m,\mu_0]} H(\{ m_{rs}/n \}) = \inf_{\mathcal{V}[k,m,f_m,\mu_0]} H(\{ x_{rs} \}). \]  

(5.1)

The proof of (5.1) is deferred till the end of the lemma.

For the lower bound, note that

\[ \mathbb{P}_n(\{ M_{rs} \}_{r,s=1}^k \in \mathcal{V}_n[k,m,f_m,\mu_0](\epsilon)) \geq \max_{\mathcal{V}_n[k,m,f_m,\mu_0]}(\epsilon) \mathbb{P}_n(M_{rs} = m_{rs}, 1 \leq r, s \leq k) \]

\[ = \max_{\mathcal{V}_n[k,m,f_m,\mu_0]}(\epsilon) \frac{\prod_{r=1}^k m_r! \prod_{s=1}^k m_s!}{n! \prod_{r,s=1}^k m_{rs}!} \]

where the second step uses Lemma 5.3. Now, Stirling’s formula gives that there exists \( C < \infty \) such that

\[ | \log n! - n \log n + n | = 0 \text{ if } n = 0 \]

\[ = 1 \text{ if } n = 1 \]

\[ \leq C \log n \text{ if } n \geq 2, \]

and so

\[ \frac{1}{n} \log \mathbb{P}_n(\mathcal{U}[k,m,\epsilon,f_m,\mu_0]) \geq -2 \log k - \min_{\mathcal{V}_n[k,m,f_m,\mu_0]} H(\{ m_{rs}/n \}) - \frac{C_k \log n}{n} \]

for some constant \( C_k < \infty \). On taking limits along with (5.1) gives

\[ \liminf_{n \to \infty} \mathbb{P}_n(\{ M_{rs} \}_{r,s=1}^k \in \mathcal{V}_n[k,m,f_m,\mu_0](\epsilon)) \geq -2 \log k - \inf_{\mathcal{V}[k,m,f_m,\mu_0]} H(\{ x_{rs} \}), \]

which completes the proof of the lower bound.

For the upper bound note that

\[ \mathbb{P}_n(\{ M_{rs} \} \in \mathcal{V}_n[k,m,f_m,\mu_0](\epsilon)) \leq \left( \frac{n + k^2 - 1}{k^2 - 1} \right) \max_{\mathcal{V}_n[k,m,f_m,\mu_0]}(\epsilon) \mathbb{P}_n(M_{rs} = m_{rs}, 1 \leq r, s \leq k) \]

\[ \leq (n + k^2)^k \max_{\mathcal{V}_n[k,m,f_m,\mu_0]}(\epsilon) \mathbb{P}_n(M_{rs} = m_{rs}, 1 \leq r, s \leq k), \]

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

\[ \frac{1}{n} \log \mathbb{P}_n(\mathcal{U}[k,m,\epsilon,f_m,\mu_0]) \leq -2 \log k - \min_{\mathcal{V}_n[k,m,f_m,\mu_0]} H(\{ m_{rs}/n \}) + \frac{C'_k \log n}{n} \]

for some other \( C'_k < \infty \), which on taking limits gives

\[ \limsup_{n \to \infty} \mathbb{P}_n(\{ M_{rs} \}_{r,s=1}^k \in \mathcal{V}_n[k,m,f_m,\mu_0](\epsilon)) \leq -2 \log k - \inf_{\mathcal{V}[k,m,f_m,\mu_0]} H(\{ x_{rs} \}), \]

completing the proof of the lemma.
It thus remains to prove (5.1). To this effect, let \(\{m_{rs}^{(n)}\}_{r,s=1}^{k}\) denote the minimizing configuration on the l.h.s. of (5.1). Then \(\{m_{rs}^{(n)}/n\}_{r,s=1}^{k}\) is a sequence in the compact set \(\{x_{rs} \geq 0 : \sum_{r,s=1}^{k} x_{rs} = 1\}\), and any convergent subsequence converges to a point in \(\mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon)]\). Thus for any \(T\) \(\lim \inf_{n \to \infty} \min_{V_n[k,m,f_{m,\mu_{0}}(\epsilon)]} H(\{m_{rs}/n\}) \geq \inf_{\mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon)]} H(\{x_{rs}\})\), where the last equality follows from since \(H(.)\) is continuous, completing the proof of the lower bound in (5.1).

Proceeding to prove the upper bound, it suffices to prove that for any \(\{x_{rs}\}_{r,s=1}^{k} \in \mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon)]\) there exists a sequence \(\{m_{rs}^{(n)}\}_{r,s=1}^{k} \in V_n[k,m,f_{m,\mu_{0}}(\epsilon)]\) such that \(\{m_{rs}^{(n)}/n\}_{r,s=1}^{k}\) converges to \(\{x_{rs}\}_{r,s=1}^{k}\) as \(n \to \infty\). To this effect, note that there exists a measure \(\mu \in \mathcal{U}\) such that \(P_{k}(\mu) = \{x_{rs}\}_{r,s=1}^{k}\). By [HKMRS, Lemma 4.2] and [HKMRS, Lemma 5.3] there exists a sequence of permutations \(\{\sigma_{n}\}\) converging to \(\mu\), and so \(\{m_{rs}^{(n)}\}_{r,s=1}^{k} := \{M_{rs}(\sigma_{n})\}_{r,s=1}^{k} \in \mathcal{A}_{k,n}\) with \(\{m_{rs}^{(n)}/n\}_{r,s=1}^{k} \to \{x_{rs}\}_{r,s=1}^{k}\). Also the set

\[
W_{k} := \left\{ \{x_{rs}\} \in \mathbb{R}^{k^{2}} : \left| \sum_{r,s=1}^{k} (x_{rs} - P_{k}(\mu_{0})) f_{i}(a_{r},b_{s}) \right| < \epsilon, 1 \leq i \leq m \right\}
\]

is open, and since \(\{x_{rs}\}_{r,s=1}^{k} \in W_{k}\), it follows that \(\{m_{rs}^{(n)}/n\}_{r,s=1}^{k} \in W_{k}\) for all large \(n\). Since \(\mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon)] = W_{k} \cap \mathcal{A}_{k,n}\), the proof is complete.

Lemma 5.7. For any \(\mathcal{U}[k,m,\epsilon,f_{m,\mu_{0}}] \in \mathcal{U}_{0}\) (see Definition (5.4),

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{n}(\mathcal{U}[k,m,\epsilon,f_{m,\mu_{0}}]) = -\inf_{\mu \in \mathcal{U}} D(P_{k}(\mu)||P_{k}(u))
\]

where \(D\) is the Kullback-Leibler divergence, and \(P_{k}\) is as defined in (5.1).

Proof. First note that

\[
\max_{1 \leq r,s \leq k} \left| P_{k}(L(\pi))(a_{r},b_{s}) - \frac{M_{rs}}{n} \right| \leq \frac{4}{n}.
\]

Indeed, since the square \(T_{rs}\) has four boundaries each of which intersect 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. This readily gives

\[
\max_{1 \leq i \leq m} \left| P_{k}(L(\pi))(f_{i}) - \sum_{r,s=1}^{k} \frac{M_{rs}}{n} f_{i}(a_{r},b_{s}) \right| \leq \frac{4N}{n}, \quad N := \max_{1 \leq i \leq m} \max_{(x,y) \in T} |f_{i}(x,y)|.
\]

Thus for any \(\delta \in (0,\epsilon)\) and all \(n\) large enough,

\[
\mathbb{P}_{n}(\mathcal{U}[k,m,\epsilon,f_{m,\mu_{0}}]) \geq \mathbb{P}_{n}(\{M_{rs}\}_{r,s=1}^{k} \in \mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon - \delta)]
\]

Using Lemma 5.6 gives

\[
\lim \inf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{n}(\{M_{rs}\}_{r,s=1}^{k} \in \mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon - \delta)]) \geq -2 \log k - \inf_{\mathcal{V}[k,m,f_{m,\mu_{0}}(\epsilon - \delta)]} H(\{x_{rs}\}).
\]
Letting \( \delta \downarrow 0 \) gives

\[
\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n (U[k, m, \epsilon, f_m, \mu_0]) \geq -2 \log k - \inf_{\nu[k, m, \epsilon, f_m, \mu_0]} H(\{x_{rs}\}). \tag{5.2}
\]

A similar argument gives

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n (\{M_{rs}\}_{r,s=1}^k \in \mathcal{V}_n[k, m, \epsilon, f_m, \mu_0](\epsilon + \delta)) \leq -2 \log k - \inf_{\nu[k, m, \epsilon, f_m, \mu_0]} H(\{x_{rs}\}),
\]

from which letting \( \delta \downarrow 0 \) gives

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n (U[k, m, \epsilon, f_m, \mu_0]) \leq -2 \log k - \inf_{\nu[k, m, \epsilon, f_m, \mu_0]} H(\{x_{rs}\}). \tag{5.3}
\]

Combining (5.2) and (5.3) gives

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n (U[k, m, \epsilon, f_m, \mu_0]) = -2 \log k - \inf_{\nu[k, m, \epsilon, f_m, \mu_0]} H(\{x_{rs}\})
\]

since \( H(.) \) is continuous. Finally note that for any \( \mu \in U[k, m, \epsilon, f_m, \mu_0], P_k(\mu) \in \mathcal{V}[k, m, \epsilon, f_m, \mu_0](\epsilon) \), and conversely for any \( \{x_{rs}\}_{r,s=1}^k \in \mathcal{V}[k, m, \epsilon, f_m, \mu_0](\epsilon) \) there exists a measure \( \mu \) such that \( P_k(\mu) = \{x_{rs}\}_{r,s=1}^k \), implying \( \mu \in U[k, m, \epsilon, f_m, \mu_0] \). Thus noting that

\[
-2 \log k - H(\{x_{rs}\}_{r,s=1}^k) = -D(\{x_{rs}\}_{r,s=1}^k || P_k(u))
\]

the result follows.

\( \square \)

**Proof of Theorem 2.1.** Since \( U_0 \) is a base for the weak topology on \( U \), by Lemma 5.7 and [DZ, Theorem 4.1.11] it follows that \( \mathbb{P}_n \) follows a weak ldp with the rate function

\[
I(\mu) = \sup_{U[k, m, \epsilon, f_m, \mu_0] \in U_0} \inf_{\nu \in U[k, m, \epsilon, f_m, \mu_0]} D(P_k(\nu)||P_k(u)).
\]

Also since \( U \) is compact it follows that full ldp holds with the good rate function \( I(.) \).

It thus remains to prove that \( I(\mu) = D(\mu||u) \). For this, first note that there is a countable collection of functions \( f \in C(T) \) which determine weak convergence on \( T \). For an explicit choice, take \( f_i, j(x, y) = x^i y^j \) for \( i, j \) non-negative integers. This works because \( T \) is compact, and so convergence in moments imply weak convergence. With \( \{f_1, f_2, \ldots \} \) denoting an arbitrary enumeration of this countable collection, setting \( V_k := U_{2k, k, \frac{1}{k}, \{f_1, \ldots, f_k\}, \mu} \in U_0 \) it follows that

\[
I(\mu) \geq \liminf_{k \to \infty} \inf_{\nu \in V_k} D(P_k(\nu)||P_k(u)).
\]

Thus, fixing \( \epsilon > 0 \), for every \( k \in \mathbb{N} \) there exists \( \nu_k \in V_k \) such that \( \inf_{\nu \in V_k} D(P_k(\nu)||P_k(u)) \geq D(P_k(\nu_k)||P_k(u)) - \epsilon \). Also by construction of \( V_k \) it follows that \( \nu_k \overset{w}{\to} \mu \), which readily implies

\[
P_k(\mu_k) \overset{w}{\to} \mu.
\]

Thus lower semi continuity of \( D(.) \) gives \( I(\mu) \geq D(\mu||u) - \epsilon \), proving the lower bound.

For the upper bound note that \( U_\mu \) contains \( \mu \), and so \( I(\mu) \leq \sup_{k \geq 1} D(P_k(\mu)||P_k(u)) \). Also note that

\[
D(\mu||u) = \sup_{f \in B(T)} \left\{ \int_{T} f d\mu - \log \int_{T} e^f d\mu \right\}, \quad D(P_k(\mu)||P_k(u)) = \sup_{f \in B_k(T)} \left\{ \int_{T} f d\mu - \log \int_{T} e^f d\mu \right\},
\]

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

\[ \square \]