Electronic Journal of Probability

Vol. 14 (2009), Paper no. 92, pages 2636–2656.

Journal URL
http://www.math.washington.edu/~ejpecp/

Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices*

Hanna Döring Peter Eichelsbacher
Fakultät für Mathematik
Ruhr-Universität Bochum
D-44780 Bochum
Germany
hanna.doering@ruhr-uni-bochum.de
peter.eichelsbacher@ruhr-uni-bochum.de

Abstract

We prove the moderate deviation principle for subgraph count statistics of Erdős-Rényi random graphs. This is equivalent in showing the moderate deviation principle for the trace of a power of a Bernoulli random matrix. It is done via an estimation of the log-Laplace transform and the Gärtner-Ellis theorem. We obtain upper bounds on the upper tail probabilities of the number of occurrences of small subgraphs. The method of proof is used to show supplemental moderate deviation principles for a class of symmetric statistics, including non-degenerate U-statistics with independent or Markovian entries.

Key words: moderate deviations, random graphs, concentration inequalities, U-statistics, Markov chains, random matrices.

AMS 2000 Subject Classification: Primary 60F10, 05C80; Secondary: 62G20, 15A52, 60F05.

Submitted to EJP on January 20, 2009, final version accepted September 20, 2009.

*The first author gratefully acknowledges the support by *Studienstiftung des deutschen Volkes*. Both authors have been supported by Deutsche Forschungsgemeinschaft via SFB/TR 12.

1 Introduction

1.1 Subgraph-count statistics

Consider an Erdős-Rényi random graph with n vertices, where for all $\binom{n}{2}$ different pairs of vertices the existence of an edge is decided by an independent Bernoulli experiment with probability p. For each $i \in \{1, \dots, \binom{n}{2}\}$, let X_i be the random variable determining if the edge e_i is present, i.e. $P(X_i = 1) = 1 - P(X_i = 0) = p(n) =: p$. The following statistic counts the number of subgraphs isomorphic to a fixed graph G with k edges and l vertices

$$W = \sum_{1 \le \kappa_1 < \dots < \kappa_k \le \binom{n}{2}} 1_{\{(e_{\kappa_1}, \dots, e_{\kappa_k}) \sim G\}} \left(\prod_{i=1}^{k} X_{\kappa_i} \right) .$$

Here $(e_{\kappa_1}, \dots, e_{\kappa_k})$ denotes the graph with edges $e_{\kappa_1}, \dots, e_{\kappa_k}$ present and $A \sim G$ denotes the fact that the subgraph A of the complete graph is isomorphic to G. We assume G to be a graph without isolated vertices and to consist of $l \ge 3$ vertices and $k \ge 2$ edges. Let the constant $a := \mathrm{aut(G)}$ denote the order of the automorphism group of G. The number of copies of G in K_n, the complete graph with n vertices and $\binom{n}{2}$ edges, is given by $\binom{n}{l} l!/a$ and the expectation of W is equal to

$$\mathbb{E}[W] = \frac{\binom{n}{l} l!}{a} p^k = \mathcal{O}(n^l p^k) .$$

It is easy to see that $P(W > 0) = o(1)$ if $p \ll n^{-l/k}$. Moreover, for the graph property that G is a subgraph, the probability that a random graph possesses it jumps from 0 to 1 at the threshold probability $n^{-1/m(G)}$, where

$$m(G) = \max \left\{ \frac{e_H}{v_H} : H \subseteq G, v_H > 0 \right\},$$

e_H, v_H denote the number of edges and vertices of $H \subseteq G$, respectively, see [JŁR00].

Limiting Poisson and normal distributions for subgraph counts were studied for probability functions $p = p(n)$. For G be an arbitrary graph, Ruciński proved in [Ruc88] that W is Poisson convergent if and only if

$$np^{d(G)} \xrightarrow{n\to\infty} 0 \quad \text{or} \quad n^{\beta(G)} p \xrightarrow{n\to\infty} 0 .$$

Here $d(G)$ denotes the density of the graph G and

$$\beta(G) := \max \left\{ \frac{v_G - v_H}{e_G - e_H} : H \subset G \right\}.$$

Consider

$$c_{n,p} := \binom{n-2}{l-2} \frac{2k}{a} (l-2)! \sqrt{\binom{n}{2} p(1-p)} p^{k-1} \tag{1.1}$$

and

$$Z := \frac{W - \mathbb{E}(W)}{c_{n,p}} = \frac{\sum_{1 \le \kappa_1 < \dots < \kappa_k \le \binom{n}{2}} 1_{\{(e_{\kappa_1}, \dots, e_{\kappa_k}) \sim G\}} \left(\prod_{i=1}^{k} X_{\kappa_i} - p^k \right)}{\binom{n-2}{l-2} \frac{2k}{a} (l-2)! \sqrt{\binom{n}{2} p(1-p)} p^{k-1}} . \tag{1.2}$$

Z has asymptotic standard normal distribution, if $np^{k-1} \overset{n\to\infty}{\longrightarrow} \infty$ and $n^2(1-p) \overset{n\to\infty}{\longrightarrow} \infty$, see Nowicki, Wierman, [NW88]. For G be an arbitrary graph with at least one edge, Ruciński proved in [Ruc88] that $\frac{W-\mathbb{E}(W)}{\sqrt{\mathbb{V}(W)}}$ converges in distribution to a standard normal distribution if and only if

$$np^{m(G)} \overset{n\to\infty}{\longrightarrow} \infty \quad \text{and} \quad n^2(1-p) \overset{n\to\infty}{\longrightarrow} \infty\,. \tag{1.3}$$

Here and in the following $\mathbb{V}$ denotes the variance of the corresponding random variable. Ruciński closed the book proving asymptotic normality in applying the method of moments. One may wonder about the normalization (1.1) used in [NW88]. The subgraph count W is a sum of dependent random variables, for which the exact calculation of the variance is tedious. In [NW88], the authors approximated W by a projection of W, which is a sum of independent random variables. For this sum the variance calculation is elementary, proving the denominator (1.1) in the definition of Z. The asymptotic behaviour of the variance of W for any $p = p(n)$ is summarized in Section 2 in [Ruc88]. The method of martingale differences used by Catoni in [Cat03] enables on the conditions $np^{3(k-\frac{1}{2})} \overset{n\to\infty}{\longrightarrow} \infty$ and $n^2(1-p) \overset{n\to\infty}{\longrightarrow} \infty$ to give an alternative proof of the central limit theorem, see remark 4.2.

A common feature is to prove large and moderate deviations, namely, the asymptotic computation of small probabilities on an exponential scale. Considering the moderate scale is the interest in the transition from a result of convergence in distribution like a central limit theorem-scaling to the large deviations scaling. Interesting enough proving that the subgraph count random variable W satisfies a large or a moderate deviation principle is an unsolved problem up to now. In [CD09] Chatterjee and Dey proved a large deviation principle for the triangle count statistic, but under the additional assumption that p is fixed and $p > 0.31$, as well as similar results with fixed probability for general subgraph counts. The main goal of this paper is to prove a moderate deviation principle for the rescaled Z, filling a substantial gap in the literature on asymptotic subgraph count distributions, see Theorem 1.1. Before we recall the definition of a moderate deviation principle and state our result, let us remark, that exponentially small probabilities have been studied extensively in the literature. A famous upper bound for *lower tails* was proven by Janson [Jan90], applying the FKG-inequality. This inequality leads to good upper bounds for the probability of nonexistence $W = 0$. Upper bounds for *upper tails* were derived by Vu [Vu01], Kim and Vu [KV04] and recently by Janson, Oleskiewicz and Ruciński [JOR04] and in [JR04] by Janson and Ruciński. A comparison of seven techniques proving bounds for *the infamous* upper tail can be found in [JR02]. In Theorem 1.3 we also obtain upper bounds on the upper tail probabilities of W.

Let us recall the definition of the large deviation principle (LDP). A sequence of probability measures $\{(\mu_n), n \in \mathbb{N}\}$ on a topological space $\mathscr{X}$ equipped with a σ-field $\mathscr{B}$ is said to satisfy the LDP with speed $s_n \nearrow \infty$ and good rate function $I(\cdot)$ if the level sets $\{x : I(x) \le \alpha\}$ are compact for all $\alpha \in [0,\infty)$ and for all $\Gamma \in \mathscr{B}$ the lower bound

$$\liminf_{n\to\infty} \frac{1}{s_n} \log \mu_n(\Gamma) \ge - \inf_{x\in \mathrm{int}(\Gamma)} I(x)$$

and the upper bound

$$\limsup_{n\to\infty} \frac{1}{s_n} \log \mu_n(\Gamma) \le - \inf_{x\in \mathrm{cl}(\Gamma)} I(x)$$

hold. Here $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ respectively. We say a sequence of random variables satisfies the LDP when the sequence of measures induced by these variables

satisfies the LDP. Formally the moderate deviation principle is nothing else but the LDP. However, we will speak about the moderate deviation principle (MDP) for a sequence of random variables, whenever the scaling of the corresponding random variables is between that of an ordinary Law of Large Numbers and that of a Central Limit Theorem.

In the following, we state one of our main results, the moderate deviation principle for the rescaled subgraph count statistic W when p is fixed, and when the sequence $p(n)$ converges to 0 or 1 sufficiently slow.

THEOREM **1.1.** *Let G be a fixed graph without isolated vertices, consisting of $k \geq 2$ edges and $l \geq 3$ vertices. The sequence $(\beta_n)_n$ is assumed to be increasing with*

$$n^{l-1}p^{k-1}\sqrt{p(1-p)} \ll \beta_n \ll n^l \left(p^{k-1}\sqrt{p(1-p)}\right)^4 . \tag{1.4}$$

Then the sequence $(S_n)_n$ of subgraph count statistics

$$S := S_n := \frac{1}{\beta_n} \sum_{1\leq\kappa_1<\cdots<\kappa_k\leq\binom{n}{2}} 1_{\{(e_{\kappa_1},\ldots,e_{\kappa_k})\sim G\}} \left(\prod_{i=1}^{k} X_{\kappa_i} - p^k\right)$$

satisfies the moderate deviation principle with speed

$$s_n = \frac{\left(\frac{2k}{a}(l-2)!\right)^2 \beta_n^2}{c_{n,p}^2} = \frac{1}{\binom{n-2}{l-2}^2\binom{n}{2}} \frac{1}{p^{2k-1}(1-p)} \beta_n^2 \tag{1.5}$$

and rate function I defined by

$$I(x) = \frac{x^2}{2\left(\frac{2k}{a}(l-2)!\right)^2} . \tag{1.6}$$

REMARKS 1.2.

1. Using $\binom{n-2}{l-2}^2\binom{n}{2} \leq n^{2(l-1)}$, we obtain $s_n \geq \left(\frac{\beta_n}{n^{l-1}p^{k-1}\sqrt{p(1-p)}}\right)^2$; therefore the condition
$$n^{l-1}p^{k-1}\sqrt{p(1-p)} \ll \beta_n$$
implies that s_n is growing to infinity as $n \to \infty$ and hence is a speed.

2. If we choose β_n such that $\beta_n \ll n^l \left(p^{k-1}\sqrt{p(1-p)}\right)^4$ and using the fact that s_n is a speed implies that
$$n^2 p^{6k-3}(1-p)^3 \overset{n\to\infty}{\longrightarrow} \infty . \tag{1.7}$$
This is a necessary but not a sufficient condition on (1.4).

Additionally the approach to prove Theorem 1.1 yields a central limit theorem for $Z = \frac{W-\mathbb{E}W}{c_{n,p}}$, see remark 4.2, and a concentration inequality for $W - \mathbb{E}W$:

THEOREM **1.3.** *Let G be a fixed graph without isolated vertices, consisting of $k \geq 2$ edges and $l \geq 3$ vertices and let W be the number of copies of G. Then for every $\varepsilon > 0$*

$$P(W - \mathbb{E}W \geq \varepsilon\mathbb{E}W) \leq \exp\left(-\frac{const.\varepsilon^2 n^{2l} p^{2k}}{n^{2l-2}p^{2k-1}(1-p) + const.\varepsilon n^{2l-2}p^{1-k}(1-p)^{-1}}\right) ,$$

where $const.$ denote constants depending on l and k only.

We will give a proof of Theorem 1.1 and Theorem 1.3 in the end of section 4.

REMARK 1.4. Let us consider the example of counting triangles: $l = k = 3$, $a = 6$. The necessary condition (1.7) of the moderate deviation principle turns to

$$n^2 p^{15} \longrightarrow \infty \quad \text{and} \quad n^2(1-p)^3 \longrightarrow \infty \quad \text{as } n \to \infty\,.$$

This can be compared to the expectedly weaker necessary and sufficient condition for the central limit theorem for Z in [Ruc88]:

$$np \longrightarrow \infty \quad \text{and} \quad n^2(1-p) \longrightarrow \infty \quad \text{as } n \to \infty.$$

The concentration inequality in Theorem 1.3 for triangles turns to

$$P(W - \mathbb{E}W \geq \varepsilon \mathbb{E}W) \leq \exp\left(-\frac{const.\varepsilon^2 n^6 p^6}{n^4 p^5(1-p) + const.\varepsilon n^4 p^{-2}(1-p)^{-1}}\right) \quad \forall \varepsilon > 0\,.$$

Kim and Vu showed in [KV04] for all $0 < \varepsilon \leq 0.1$ and for $p \geq \frac{1}{n}\log n$, that

$$P\left(\frac{W - \mathbb{E}W}{\varepsilon p^3 n^3} \geq 1\right) \leq e^{-\Theta(p^2 n^2)}\,.$$

As we will see in the proof of Theorem 1.3, the bound for $d(n)$ in (1.12) leads to an additional term of order $n^2 p^8$. Hence in general our bounds are not optimal. Optimal bounds were obtained only for some subgraphs. Our concentration inequality can be compared with the bounds in [JR02], which we leave to the reader.

1.2 Bernoulli random matrices

Theorem 1.1 can be reformulated as the moderate deviation principle for traces of a power of a Bernoulli random matrix.

THEOREM **1.5.** *Let $X = (X_{ij})_{i,j}$ be a symmetric $n \times n$-matrix of independent real-valued random variables, Bernoulli-distributed with probability*

$$P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p(n),\ i < j$$

and $P(X_{ii} = 0) = 1$, $i = 1, \ldots, n$. Consider for any fixed $k \geq 3$ the trace of the matrix to the power k

$$Tr(X^k) = \sum_{i_1,\ldots,i_k=1}^{n} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}\,. \tag{1.8}$$

Note that $Tr(X^k) = 2W$, for W counting circles of length k in a random graph. We obtain that the sequence $(T_n)_n$ with

$$T_n := \frac{Tr(X^k) - \mathbb{E}[Tr(X^k)]}{2\beta_n} \tag{1.9}$$

satisfies the moderate deviation principle for any β_n satisfying (1.4) *with $l = k$ and with rate function* (1.6) *with $l = k$ and $a = 2k$:*

$$I(x) = \frac{x^2}{2\left((k-2)!\right)^2}\,. \tag{1.10}$$

REMARK 1.6. The following is a famous open problem in random matrix theory: Consider X to be a symmetric $n \times n$ matrix with entries X_{ij} $(i \le j)$ being i.i.d., satisfying some exponential integrability. The question is to prove for any fixed $k \ge 3$ the LDP for

$$\frac{1}{n^k}\mathrm{Tr}(X^k)$$

and the MDP for

$$\frac{1}{\beta_n(k)}\big(\mathrm{Tr}(X^k) - \mathbb{E}[\mathrm{Tr}(X^k)]\big) \tag{1.11}$$

for a properly chosen sequence $\beta_n(k)$. For $k = 1$ the LDP in question immediately follows from Cramér's theorem (see [DZ98, Theorem 2.2.3]), since

$$\frac{1}{n}\mathrm{Tr}(X) = \frac{1}{n}\sum_{i=1}^{n} X_{ii}.$$

For $k = 2$, notice that

$$\frac{1}{n^2}\mathrm{Tr}(X^2) = \frac{2}{n^2}\sum_{i<j} X_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n} X_{ii}^2 =: A_n + B_n.$$

By Cramér's theorem we know that $(\tilde{A}_n)_n$ with $\tilde{A}_n := \frac{1}{\binom{n}{2}}\sum_{i<j} X_{ij}^2$ satisfies the LDP, and by Chebychev's inequality we obtain for any $\varepsilon > 0$

$$\limsup_{n\to\infty} \frac{1}{n}\log P(|B_n| \ge \varepsilon) = -\infty.$$

Hence $(A_n)_n$ and $(\frac{1}{n^2}\mathrm{Tr}(X^2))_n$ are exponentially equivalent (see [DZ98, Definition 4.2.10]). Moreover $(A_n)_n$ and $(\tilde{A}_n)_n$ are exponentially equivalent, since Chebychev's inequality leads to

$$\limsup_{n\to\infty} \frac{1}{n}\log P(|A_n - \tilde{A}_n| > \varepsilon) = \limsup_{n\to\infty} \frac{1}{n}\log P\left(|\sum_{i<j} X_{ij}^2| \ge \varepsilon \frac{n^2(n-1)}{2}\right) = -\infty.$$

Applying Theorem 4.2.13 in [DZ98], we obtain the LDP for $(1/n^2\mathrm{Tr}(X^2))_n$ under exponential integrability. For $k \ge 3$, proving the LDP for $(1/n^k\mathrm{Tr}(X^k))_n$ is open, even in the Bernoulli case. For Gaussian entries X_{ij} with mean 0 and variance $1/n$, the LDP for the sequence of empirical measures of the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, e.g.

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

has been established by Ben Arous and Guionnet in [BAG97]. Although one has the representation

$$\frac{1}{n^k}\mathrm{Tr}(X^k) = \frac{1}{n^{k/2}}\mathrm{Tr}\left(\frac{X}{\sqrt{n}}\right)^k = \frac{1}{n^{k/2}}\sum_{i=1}^{n} \lambda_i^k,$$

the LDP cannot be deduced from the LDP of the empirical measure by the contraction principle [DZ98, Theorem 4.2.1], because $x \to x^k$ is not bounded in this case.

REMARK 1.7. Theorem 1.5 told us that in the case of Bernoulli random variables X_{ij}, the MDP for (1.11) holds for any $k \geq 3$. For $k = 1$ and $k = 2$, the MDP for (1.11) holds for arbitrary i.i.d. entries X_{ij} satisfying some exponential integrability: For $k = 1$ we choose $\beta_n(1) := a_n$ with a_n any sequence with $\lim_{n\to\infty} \frac{\sqrt{n}}{a_n} = 0$ and $\lim_{n\to\infty} \frac{n}{a_n} = \infty$. For

$$\frac{1}{a_n} \sum_{i=1}^{n} (X_{ii} - \mathbb{E}(X_{ii}))$$

the MDP holds with rate $x^2/(2\mathbb{V}(X_{11}))$ and speed a_n^2/n, see Theorem 3.7.1 in [DZ98]. In the case of Bernoulli random variables, we choose $\beta_n(1) = a_n$ with $(a_n)_n$ any sequence with

$$\lim_{n\to\infty} \frac{\sqrt{np(1-p)}}{a_n} = 0 \text{ and } \lim_{n\to\infty} \frac{n\sqrt{p(1-p)}}{a_n} = \infty$$

and $p = p(n)$. Now $(\frac{1}{a_n} \sum_{i=1}^{n} (X_{ii} - \mathbb{E}(X_{ii})))_n$ satisfies the MDP with rate function $x^2/2$ and speed

$$\frac{a_n^2}{np(n)(1-p(n))}.$$

Hence, in this case $p(n)$ has to fulfill the condition $n^2 p(n)(1-p(n)) \to \infty$.

For $k = 2$, we choose $\beta_n(2) = a_n$ with a_n being any sequence with $\lim_{n\to\infty} \frac{n}{a_n} = 0$ and $\lim_{n\to\infty} \frac{n^2}{a_n} = \infty$. Applying Chebychev's inequality and exponential equivalence arguments similar as in Remark 1.6, we obtain the MDP for

$$\frac{1}{a_n} \sum_{i,j=1}^{n} (X_{ij}^2 - \mathbb{E}(X_{ij}^2))$$

with rate $x^2/(2\mathbb{V}(X_{11}))$ and speed a_n^2/n^2.The case of Bernoulli random variables can be obtained in a similar way.

REMARK 1.8. For $k \geq 3$ we obtain the MDP with $\beta_n = \beta_n(k)$ such that

$$n^{k-1} p(n)^{k-1} \sqrt{p(n)(1-p(n))} \ll \beta_n \ll n^k \big(p(n)^{k-1} \sqrt{p(n)(1-p(n))}\big)^4.$$

Considering a fixed p, the range of β_n is what we should expect: $n^{k-1} \ll \beta_n \ll n^k$. But we also obtain the MDP for functions $p(n)$. In random matrix theory, Wigner 1959 analysed Bernoulli random matrices in Nuclear Physics. Interestingly enough, the moderate deviation principle for the empirical mean of the eigenvalues of a random matrix is known only for symmetric matrices with Gaussian entries and for non-centered Gaussian entries, respectively, see [DGZ03]. The proofs depend on the existence of an explicit formula for the joint distribution of the eigenvalues or on corresponding matrix-valued stochastic processes.

1.3 Symmetric Statistics

On the way of proving Theorem 1.1, we will apply a nice result of Catoni [Cat03, Theorem 1.1]. Doing so, we recognized, that Catoni's approach lead us to a general approach proving the moderate deviation principle for a rich class of statistics, which -without loss of generality- can be assumed

to be *symmetric statistics*. Let us make this more precise. In [Cat03], non-asymptotic bounds of the log-Laplace transform of a function f of $k(n)$ random variables $X := (X_1, \dots, X_{k(n)})$ lead to concentration inequalities. These inequalities can be obtained for independent random variables or for Markov chains. It is assumed in [Cat03] that the *partial finite differences of order one and two* of f are suitably bounded. The line of proof is a combination of a martingale difference approach and a Gibbs measure philosophy.

Let $(\Omega, \mathscr{A})$ be the product of measurable spaces $\otimes_{i=1}^{k(n)}(\mathscr{X}_i, \mathscr{B}_i)$ and $\mathbb{P} = \otimes_{i=1}^{k(n)}\mu_i$ be a product probability measure on $(\Omega, \mathscr{A})$. Let $X_1, \dots, X_{k(n)}$ take its values in $(\Omega, \mathscr{A})$ and assume that $(X_1, \dots, X_{k(n)})$ is the canonical process. Let $(Y_1, \dots, Y_{k(n)})$ be an independent copy of $X := (X_1, \dots, X_{k(n)})$ such that Y_i is distributed according to μ_i, $i = 1, \dots, k(n)$. The function $f : \Omega \to \mathbb{R}$ is assumed to be *bounded and measurable*.
Let $\Delta_i f(x_1^{k(n)}; y_i)$ denote the *partial difference of order one* of f defined by

$$\Delta_i f(x_1^{k(n)}; y_i) := f(x_1, \dots, x_{k(n)}) - f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_{k(n)}),$$

where $x_1^{k(n)} := (x_1, \dots, x_{k(n)}) \in \Omega$ and $y_i \in \mathscr{X}_i$. Analogously we define for $j < i$ and $y_j \in \mathscr{X}_j$ the *partial difference of order two*

$$\begin{aligned}\Delta_i \Delta_j f(x_1^{k(n)}; y_j, y_i) :=& \Delta_i f(x_1^{k(n)}; y_i) - f(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_{k(n)}) \\ &+ f(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_{k(n)}).\end{aligned}$$

Now we can state our main theorem. If the random variables are independent and if the partial finite differences of the first and second order of f are suitably bounded, then f, properly rescaled, satisfies the MDP:

THEOREM **1.9.** *In the above setting assume that the random variables in X are independent. Define $d(n)$ by*

$$d(n) := \sum_{i=1}^{k(n)} |\Delta_i f(X_1^{k(n)}; Y_i)|^2 \left(\frac{1}{3}|\Delta_i f(X_1^{k(n)}; Y_i)| + \frac{1}{4}\sum_{j=1}^{i-1} |\Delta_i \Delta_j f(X_1^{k(n)}; Y_j, Y_i)| \right). \tag{1.12}$$

Note that here and in the following $|\cdot|$ of a random term denotes the maximal value of the term for all $\omega \in \Omega$. Moreover let there exist two sequences $(s_n)_n$ and $(t_n)_n$ such that

1. $\frac{s_n^2}{t_n^3} d(n) \overset{n\to\infty}{\longrightarrow} 0$ *and*
2. $\frac{s_n}{t_n^2} \mathbb{V} f(X) \overset{n\to\infty}{\longrightarrow} C > 0$ *for the variance of f.*

Then the sequence of random variables

$$\left(\frac{f(X) - \mathbb{E}[f(X)]}{t_n} \right)_n$$

satisfies the moderate deviation principle with speed s_n and rate function $\frac{x^2}{2C}$.

In Section 2 we are going to prove Theorem 1.9 via the Gärtner-Ellis theorem. In [Cat03] an inequality has been proved which allows to relate the logarithm of a Laplace transform with the expectation and the variance of the observed random variable. Catoni proves a similar result for the logarithm of a Laplace transform of random variables with Markovian dependence. One can find a different $d(n)$ in [Cat03, Theorem 3.1]. To simplify notations we did not generalize Theorem 1.9, but the proof can be adopted immediately. In Section 3 we obtain moderate deviations for several symmetric statistics, including the sample mean and U-statistics with independent and Markovian entries. In Section 4 we proof Theorem 1.1 and 1.3.

2 Moderate Deviations via Laplace Transforms

Theorem 1.9 is an application of the following theorem:

THEOREM **2.1.** *(Catoni, 2003)*
In the setting of Theorem 1.9, assuming that the random variables in X are independent, one obtains for all $s \in \mathbb{R}_+$,

$$\left|\log \mathbb{E}\exp\left(sf(X)\right) - s\mathbb{E}\left[f(X)\right] - \frac{s^2}{2}\mathbb{V}f(X)\right| \leq s^3 d(n) \tag{2.13}$$
$$= \sum_{i=1}^{k(n)} \frac{s^3}{3} |\Delta_i f(X_1^{k(n)}; Y_i)|^3 + \sum_{i=1}^{k(n)} \sum_{j=1}^{i-1} \frac{s^3}{4} |\Delta_i f(X_1^{k(n)}; Y_i)|^2 |\Delta_i \Delta_j f(X_1^{k(n)}; Y_j, Y_i)| \,.$$

Proof of Theorem 2.1. We decompose $f(X)$ into martingale differences

$$F_i(f(X)) = \mathbb{E}\left[f(X)\middle|X_1,\ldots,X_i\right] - \mathbb{E}\left[f(X)\middle|X_1,\ldots,X_{i-1}\right] \quad \text{, for all } i \in \{1,\ldots,k(n)\} \,.$$

The variance can be represented by $\mathbb{V}f(X) = \sum_{i=1}^{k(n)} \mathbb{E}\left[\left(F_i(f(X))\right)^2\right]$.

Catoni uses the triangle inequality and compares the two terms $\log \mathbb{E}e^{sf(X)-s\mathbb{E}[f(X)]}$ and $\frac{s^2}{2}\mathbb{V}f(X)$ to the above representation of the variance with respect to the Gibbs measure with density

$$dP_W := \frac{e^W}{\mathbb{E}[e^W]}\, dP,$$

where W is a bounded measurable function of $(X_1,\ldots,X_{k(n)})$. We denote an expectation due to this Gibbs measure by $\mathbb{E}_W$, e.g.

$$\mathbb{E}_W[X] := \frac{\mathbb{E}[X\exp(W)]}{\mathbb{E}[\exp(W)]} \,.$$

On the one hand Catoni bounds the difference

$$\left|\log \mathbb{E}e^{sf(X)-s\mathbb{E}[f(X)]} - \frac{s^2}{2}\sum_{i=1}^{k(n)} \mathbb{E}_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\ldots,X_{i-1}\right]}\left[\left(F_i\left(f(X)\right)\right)^2\right]\right|$$

via partial integration:

$$\begin{aligned}
&\left|\log \mathbb{E}e^{s\left(f(X)-\mathbb{E}[f(X)]\right)} - \frac{s^2}{2}\sum_{i=1}^{k(n)} \mathbb{E}_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]}\left[F_i^2(f(X))\right]\right| \\
= &\left|\sum_{i=1}^{k(n)}\int_0^s \frac{(s-\alpha)^2}{2}\mathbb{M}^3_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]+\alpha F_i(f(X))}[F_i(f(X))]d\alpha\right|,
\end{aligned}$$

where $\mathbb{M}^3_U[X] := \mathbb{E}_U\left[\left(X-\mathbb{E}_U[X]\right)^3\right]$ for a bounded measurable function U of $(X_1,\dots,X_{k(n)})$. Moreover

$$\begin{aligned}
&\left|\sum_{i=1}^{k(n)}\int_0^s \frac{(s-\alpha)^2}{2}\mathbb{M}^3_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]+\alpha F_i(f(X))}[F_i(f(X))]d\alpha\right| \\
\le &\left|\sum_{i=1}^{k(n)}||F_i(f(X))||^3_\infty\int_0^s(s-\alpha)^2d\alpha\right| \le \sum_{i=1}^{k(n)}\frac{s^3}{3}|\Delta_i f(X_1^{k(n)};Y_i)|^3\,.
\end{aligned}$$

On the other hand he uses the following calculation:

$$\begin{aligned}
&\left|\frac{s^2}{2}\sum_{i=1}^{k(n)}\mathbb{E}_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]}\left[\left(F_i(f(X))\right)^2\right] - \frac{s^2}{2}\mathbb{V}f(X)\right| \\
= &\left|\frac{s^2}{2}\sum_{i=1}^{k(n)}\mathbb{E}_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]}\left[\left(F_i(f(X))\right)^2\right] - \frac{s^2}{2}\sum_{i=1}^{k(n)}\mathbb{E}\left[\left(F_i(f(X))\right)^2\right]\right| \\
= &\frac{s^2}{2}\sum_{i=1}^{k(n)}\sum_{j=1}^{i-1}\mathbb{E}_{s\mathbb{E}\left[f(X)-\mathbb{E}[f(X)]\middle|X_1,\dots,X_{i-1}\right]}\left[F_j\left(\left(F_i(f(X))\right)^2\right)\right] \\
\le &\frac{s^2}{2}\sum_{i=1}^{k(n)}\sum_{j=1}^{i-1}\int_0^s\sqrt{\mathbb{E}^{\mathscr{G}_j}_{\alpha G_{j,i-1}}[F_j\left(F_i^2(f(X))\right)^2]\,\mathbb{E}^{\mathscr{G}_j}_{\alpha G_{j,i-1}}[W^2]}d\alpha
\end{aligned}$$

applying the Cauchy-Schwartz inequality and the notation

$$\begin{aligned}
&E^{\mathscr{G}_j}[\cdot] := \mathbb{E}[\cdot|X_1,\dots,X_{j-1},X_{j+1},\dots,X_{k(n)}] \quad \text{and} \\
&W = G_{j,i-1} - \mathbb{E}^{\mathscr{G}_j}_{\alpha G_{j,i-1}}[G_{j,i-1}], \\
&\text{where } G_{j,i-1} = \mathbb{E}\left[f(X)\middle|X_1,\dots,X_{i-1}\right] - \mathbb{E}^{\mathscr{G}_j}\left[\mathbb{E}\left[f(X)\middle|X_1,\dots,X_{i-1}\right]\right].
\end{aligned}$$

As you can see in [Cat03] $F_j\left(F_i^2(f(X))\right)^2$ and W can be estimated in terms of $\Delta_i f(X)$ and $\Delta_i\Delta_j f(X)$, independently of the variable of integration α. This leads to the inequality stated in Theorem 2.1. □

Proof of Theorem 1.9. To use the Gärtner-Ellis theorem (see [DZ98, Theorem2.3.6]) we have to calculate the limit of

$$\frac{1}{s_n}\log\mathbb{E}\exp\left(\lambda s_n\frac{f(X)-\mathbb{E}[f(X)]}{t_n}\right) = \frac{1}{s_n}\left(\log\mathbb{E}\exp\left(\frac{\lambda s_n}{t_n}f(X)\right) - \frac{\lambda s_n}{t_n}\mathbb{E}[f(X)]\right) \tag{2.14}$$

for $\lambda \in \mathbb{R}$. We apply Theorem 2.1 for $s = \dfrac{\lambda s_n}{t_n}$ and $\lambda > 0$. The right hand side of the inequality (2.13) converges to zero for large n:

$$\frac{1}{s_n} s^3 d(n) = \lambda^3 \frac{s_n^2}{t_n^3} d(n) \overset{n\to\infty}{\longrightarrow} 0 \tag{2.15}$$

as assumed in condition (1). Applying (2.13) this leads to the limit

$$\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{s_n} \log \mathbb{E} \exp\left(\lambda s_n \frac{f(X) - \mathbb{E}[f(X)]}{t_n}\right) = \lim_{n\to\infty} \frac{1}{s_n} \frac{\lambda^2 s_n^2}{2t_n^2} \mathbb{V} f(X) = \frac{\lambda^2}{2} C, \tag{2.16}$$

where the last equality follows from condition (2). Λ is finite and differentiable. The same calculation is true for $-f$ and consequently (2.16) holds for all $\lambda \in \mathbb{R}$. Hence we are able to apply the Gärtner-Ellis theorem. This proves the moderate deviation principle of $\left(\frac{f(X)-\mathbb{E}[f(X)]}{t_n}\right)_n$ with speed s_n and rate function

$$I(x) = \sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \frac{\lambda^2}{2} C\right\} = \frac{x^2}{2C}.$$

□

3 Moderate Deviations for Non-degenerate U-statistics

In this section we show three applications of Theorem 1.9. We start with the simplest case:

3.1 sample mean

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with values in a compact set $[-r, r], r > 0$ fix, and positive variance as well as $Y_1, \ldots, Y_n$ independent copies. To apply Theorem 1.9 for $f(X) = \frac{1}{\sqrt{n}} \sum_{m=1}^n X_m$ the partial differences of f have to tend to zero fast enough for n to infinity:

$$|\Delta_i f(X_1^n; Y_i)| = \frac{1}{\sqrt{n}} |X_i - Y_i| \leq \frac{2r}{\sqrt{n}} \tag{3.17}$$

$$\Delta_i \Delta_j f(X_1^n; Y_j, Y_i) = 0 \tag{3.18}$$

Let a_n be a sequence with $\lim_{n\to\infty} \frac{\sqrt{n}}{a_n} = 0$ and $\lim_{n\to\infty} \frac{n}{a_n} = \infty$. For $t_n = \frac{a_n}{\sqrt{n}}$ and $s_n = \frac{a_n^2}{n}$ the conditions of Theorem 1.9 are satisfied:

1. $\dfrac{s_n^2}{t_n^3} d(n) \leq \dfrac{a_n}{\sqrt{n}} \dfrac{4r^2}{n} \sum_{m=1}^n \dfrac{2r}{3\sqrt{n}} = \dfrac{a_n}{n} \dfrac{8r^3}{3}$. Because $d(n)$ is positive this implies $\lim\limits_{n\to\infty} \dfrac{s_n^2}{t_n^3} d(n) = 0$.

2. $\dfrac{s_n}{t_n^2} \mathbb{V} f(X) = \mathbb{V}\left(\dfrac{1}{\sqrt{n}} \sum_{m=1}^n X_m\right) = \mathbb{V}(X_1)$.

The application of Theorem 1.9 proves the MDP for $\frac{1}{a_n}\left(\sum_{m=1}^{n} X_m - n\mathbb{E}X_1\right)_n$ with speed s_n and rate function $I(x) = \frac{x^2}{2\mathbb{V}X_1}$. This result is well known, see for example [DZ98], Theorem 3.7.1, and references therein. The MDP can be proved under *local exponential moment* conditions on X_1: $\mathbb{E}(\exp(\lambda X_1)) < \infty$ for a $\lambda > 0$. In [Cat03], the bounds of the log-Laplace transformation are obtained under exponential moment conditions. Applying this result, we would be able to obtain the MDP under exponential moment conditions, but this is not the focus of this paper.

3.2 non-degenerate U-statistics with independent entries

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with values in a measurable space $\mathcal{X}$. For a measurable and symmetric function $h : \mathcal{X}^m \to \mathbb{R}$ we define

$$U_n(h) := \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}),$$

where symmetric means invariant under all permutation of its arguments. $U_n(h)$ is called a *U-statistic* with *kernel* h and *degree* m.

Define the conditional expectation for $c = 1, \ldots, m$ by

$$\begin{aligned} h_c(x_1, \ldots, x_c) &:= \mathbb{E}\big[h(x_1, \ldots, x_c, X_{c+1}, \ldots, X_m)\big] \\ &= \mathbb{E}\big[h(X_1, \ldots, X_m)\big|X_1 = x_1, \ldots, X_c = x_c\big] \end{aligned}$$

and the variances by $\sigma_c^2 := \mathbb{V}\big[h_c(X_1, \ldots, X_c)\big]$. A U-statistic is called *non-degenerate* if $\sigma_1^2 > 0$.

By the Hoeffding-decomposition (see for example [Lee90]), we know that for every symmetric function h, the U-statistic can be decomposed into a sum of degenerate U-statistics of different orders. In the degenerate case the linear term of this decomposition disappears. Eichelsbacher and Schmock showed the MDP for non-degenerate U-statistics in [ES03]; the proof used the fact that the linear term in the Hoeffding-decomposition is leading in the non-degenerate case. In this article the observed U-statistic is assumed to be of the latter case.

We show the MDP for appropriate scaled U-statistics without applying Hoeffding's decomposition. The scaled U-statistic $f := \sqrt{n}U_n(h)$ with bounded kernel h and degree 2 fulfils the inequality:

$$\begin{aligned} \Delta_k f(x_1^n; y_k) &= \frac{2\sqrt{n}}{n(n-1)}\Big(\sum_{1 \le i < j \le n} h(x_i, x_j) - \sum_{\substack{1 \le i < j \le n \\ i,j \ne k}} h(x_i, x_j) - \sum_{i=1}^{k-1} h(x_i, y_k) \\ &\quad - \sum_{j=k+1}^{n} h(y_k, x_j)\Big) \\ &= \frac{2}{\sqrt{n}(n-1)}\left(\sum_{i=1}^{k-1} h(x_i, x_k) + \sum_{j=k+1}^{n} h(x_k, x_j) - \sum_{i=1}^{k-1} h(x_i, y_k) - \sum_{j=k+1}^{n} h(y_k, x_j)\right) \\ &\le \frac{4||h||_\infty}{\sqrt{n}} \end{aligned}$$

for $k = 1,\dots,n$. Analogously one can write down all summations of the kernel h for $\Delta_m\Delta_k f(x_1^n; y_k, y_m)$. Most terms add up to zero and we get:

$$\begin{aligned}\Delta_m\Delta_k f(x_1^n; y_k, y_m) &= \frac{2\left(h(x_k,x_m) - h(y_k,x_m) - h(x_k,y_m) + h(y_k,y_m)\right)}{\sqrt{n}(n-1)} \\ &\le \frac{2}{\sqrt{n}(n-1)} 4||h||_\infty \le \frac{16||h||_\infty}{n^{3/2}}\,.\end{aligned}$$

Let a_n be a sequence with $\lim_{n\to\infty} \frac{\sqrt{n}}{a_n} = 0$ and $\lim_{n\to\infty} \frac{n}{a_n} = \infty$. The aim is the MDP for a real random variable of the kind $\frac{n}{a_n} U_n(h)$ and the speed $s_n := \frac{a_n^2}{n}$. To apply Theorem 1.9 for $f(X) = \sqrt{n}U_n(h)(X)$, s_n as above and $t_n := \frac{a_n}{\sqrt{n}}$, we obtain

1. $\frac{s_n^2}{t_n^3} d(n) \le \frac{a_n}{\sqrt{n}} \left(\frac{4||h||_\infty^3}{3\sqrt{n}} + \frac{n-1}{n^{3/2}} 8||h||_\infty^3 \right)$. The right hand side converges to 0, because $\lim_{n\to\infty} a_n/n = 0$.
2. $\frac{s_n}{t_n^2}\mathbb{V} f(X) = \frac{a_n^2}{n}\frac{n}{a_n^2}\mathbb{V}\left(\sqrt{n}U_n(h)(X)\right) \overset{n\to\infty}{\longrightarrow} 4\sigma_1^2$, see Theorem 3 in [Lee90, chapter 1.3].

The non-degeneracy of $U_n(h)$ implies that $4\sigma_1^2 > 0$.
The application of Theorem 1.9 proves:

THEOREM **3.1.** *Let $(a_n)_n \in (0,\infty)^{\mathbb{N}}$ be a sequence with $\lim_{n\to\infty} \frac{\sqrt{n}}{a_n} = 0$ and $\lim_{n\to\infty} \frac{n}{a_n} = \infty$. Then the sequence of non-degenerate and centered U-statistics $\left(\frac{n}{a_n} U_n(h)\right)_n$ with a real-valued, symmetric and bounded kernel function h satisfies the MDP with speed $s_n := \frac{a_n^2}{n}$ and good rate function*

$$I(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - 2\lambda^2\sigma_1^2\} = \frac{x^2}{8\sigma_1^2}\,.$$

REMARK 3.2. Theorem 3.1 holds, if the kernel function h depends on i and j, e.g. the U-statistic is of the form $\frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} h_{i,j}(X_i, X_j)$. One can see this in the estimation of $\Delta_i f(X)$ and $\Delta_i\Delta_j f(X)$. This is an improvement of the result in [ES03].

REMARK 3.3. We considered U-statistics with degree 2. For degree $m > 2$ we get the following estimation for the partial differences of

$$\begin{aligned} f(X) &:= \frac{1}{\sqrt{n}\binom{n}{m}} \sum_{1\le i_1<\dots<i_m\le n} h(X_{i_1},\dots,X_{i_m}) : \\ \Delta_i f(X) &\le \sqrt{n}\frac{1}{\binom{n}{m}} \binom{n-1}{m-1} 2||h||_\infty = \frac{2m}{\sqrt{n}}||h||_\infty \\ \Delta_i\Delta_j f(X) &\le \sqrt{n}\frac{1}{\binom{n}{m}} \binom{n-2}{m-2} 4||h||_\infty = \frac{4m(m-1)}{\sqrt{n}(n-1)}||h||_\infty \end{aligned}$$

and Theorem 1.9 can be applied as before.

Theorem 3.1 is proved in [ES03] in a more general context. Eichelsbacher and Schmock showed the moderate deviation principle for degenerate and non-degenerate U-statistics with a kernel function h, which is bounded or satisfies exponential moment conditions (see also [Eic98; Eic01]).

Example 1: Consider the *sample variance* $U_n^{\mathbb{V}}$, which is a U-statistic of degree 2 with kernel $h(x_1,x_2)=\frac{1}{2}(x_1-x_2)^2$. Let the random variables $X_i, i=1,\dots,n$, be restricted to take values in a compact interval. A simple calculation shows

$$\sigma_1^2=\mathbb{V}\big[h_1(X_1)\big]=\frac{1}{4}\mathbb{V}\big[(X_1-\mathbb{E}X_1)^2\big]=\frac{1}{4}\left(\mathbb{E}[(X_1-\mathbb{E}X_1)^4]-(\mathbb{V}X_1)^2\right).$$

The U-statistic is non-degenerate, if the condition $\mathbb{E}[(X_1-\mathbb{E}X_1)^4]>(\mathbb{V}X_1)^2$ is satisfied. Then $\left(\frac{n}{a_n(n-1)}\sum_{i=1}^n(X_i-\bar{X})^2\right)_n$ satisfies the MDP with speed $\frac{a_n^2}{n}$ and good rate function

$$I^{\mathbb{V}}(x)=\frac{x^2}{8\sigma_1^2}=\frac{x^2}{2\left(\mathbb{E}[(X_1-\mathbb{E}X_1)^4]-(\mathbb{V}X_1)^2\right)}.$$

In the case of independent Bernoulli random variables with $P(X_1=1)=1-P(X_1=0)=p$, $0<p<1$, $U_n^{\mathbb{V}}$ is a non-degenerate U-statistic for $p\neq\frac{1}{2}$ and the corresponding rate function is given by:

$$I^{\mathbb{V}}_{\text{bernoulli}}(x)=\frac{x^2}{2p(1-p)\big(1-4p(1-p)\big)}.$$

Example 2: The *sample second moment* is defined by the kernel function $h(x_1,x_2)=x_1x_2$. This leads to

$$\sigma_1^2=\mathbb{V}\big(h_1(X_1)\big)=\mathbb{V}\big(X_1\mathbb{E}X_1\big)=\big(\mathbb{E}X_1\big)^2\mathbb{V}X_1.$$

The condition $\sigma_1^2>0$ is satisfied, if the expectation and the variance of the observed random variables are unequal to zero. The values of the random variables have to be in a compact interval as in the example above. Under this conditions $\frac{n}{a_n}\sum_{1\le i<j\le n}X_iX_j$ satisfies the MDP with speed $\frac{a_n^2}{n}$ and good rate function

$$I^{\text{sec}}(x)=\frac{x^2}{8\sigma_1^2}=\frac{x^2}{8\big(\mathbb{E}X_1\big)^2\mathbb{V}X_1}.$$

For independent Bernoulli random variables the rate function for all $0<p<1$ is:

$$I^{\text{sec}}_{\text{bernoulli}}(x)=\frac{x^2}{8p^3(1-p)}.$$

Example 3: Wilcoxon one sample statistic Let $X_1,\dots,X_n$ be real valued, independent and identically distributed random variables with absolute continuous distribution function symmetric in zero. We prove the MDP for -properly rescaled-

$$W_n=\sum_{1\le i<j\le n}1_{\{X_i+X_j>0\}}=\binom{n}{2}U_n(h)$$

defining $h(x_1,x_2):=1_{\{x_1+x_2>0\}}$ for all $x_1,x_2\in\mathbb{R}$. Under these assumptions one can calculate $\sigma_1^2=\mathrm{Cov}\big(h(X_1,X_2),h(X_2,X_3)\big)=\frac{1}{12}$. Applying Theorem 3.1 as before we proved the MDP for the Wilcoxon one sample statistic $\frac{1}{(n-1)a_n}\left(W_n-\frac{1}{2}\binom{n}{2}\right)$ with speed $\frac{a_n^2}{n}$ and good rate function $I^W(x)=\frac{3}{2}x^2$.

3.3 non-degenerate U-statistics with Markovian entries

The moderate deviation principle in Theorem 1.9 is stated for independent random variables. Catoni showed in [Cat03], that the estimation of the logarithm of the Laplace transform can be generalized for Markov chains via a coupled process. In the following one can see, that analogously to the proof of Theorem 1.9 these results yield the moderate deviation principle.

In this section we use the notation introduced in [Cat03], Chapter 3.

Let us assume that $(X_k)_{k\in\mathbb{N}}$ is a Markov chain such that for $X := (X_1,\dots,X_n)$ the following inequalities hold

$$P\big(\tau_i > i+k \big| \mathscr{G}_i, X_i\big) \le A\rho^k \quad \forall k\in\mathbb{N} \quad a.s. \tag{3.19}$$

$$P\big(\tau_i > i+k \big| \mathscr{F}_n, \overset{i}{Y}_i\big) \le A\rho^k \quad \forall k\in\mathbb{N} \quad a.s. \tag{3.20}$$

for some positive constants A and $\rho < 1$. Here $\overset{i}{Y} := (\overset{i}{Y}_1,\dots,\overset{i}{Y}_n)$, $i = 1,\dots,n$, are n coupled stochastic processes satisfying for any i that $\overset{i}{Y}$ is equal in distribution to X. For the list of the properties of these coupled processes, see page 14 in [Cat03]. Moreover, the σ-algebra $\mathscr{G}_i$ in (3.19) is generated by $\overset{i}{Y}$, the σ-algebra $\mathscr{F}_n$ in (3.20) is generated by $(X_1,\dots,X_n)$. Finally the coupling stopping times τ_i are defined as

$$\tau_i = \inf\{k \ge i | \overset{i}{Y}_k = X_k\}.$$

Now we can state our result:

THEOREM **3.4.** *Let us assume that $(X_k)_{k\in\mathbb{N}}$ is a Markov chain such that for $X := (X_1,\dots,X_n)$ (3.19) and (3.20) hold true. Let $U_n(h)(X)$ be a non-degenerate U-statistic with bounded kernel function h and $\lim_{n\to\infty}\mathbb{V}\big(\sqrt{n}U_n(h)(X)\big) < \infty$. Then for every sequence a_n, where*

$$\lim_{n\to\infty}\frac{a_n}{n} = 0 \text{ and } \lim_{n\to\infty}\frac{n}{a_n^2} = 0\,,$$

the sequence $\big(\frac{n}{a_n}U_n(h)(X)\big)_n$ satisfies the moderate deviation principle with speed $s_n = \frac{a_n^2}{n}$ and rate function I given by

$$I(x) := \sup_{\lambda\in\mathbb{R}}\left\{\lambda x - \frac{\lambda^2}{2}\lim_{n\to\infty}\mathbb{V}\big(\sqrt{n}U_n(h)(X)\big)\right\}.$$

Proof. As for the independent case we define $f(X) := \sqrt{n}U_n(h)(X_1,\dots,X_n)$. Corollary 3.1 of [Cat03] states, that in the above situation the inequality

$$\begin{aligned}
&\left|\log\mathbb{E}\exp\big(sf(X)\big) - s\mathbb{E}\big[f(X)\big] - \frac{s^2}{2}\mathbb{V}f(X)\right| \\
\le&\frac{s^3}{\sqrt{n}}\frac{BCA^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{s}{\sqrt{n}}\right)_+^{-1} \\
&+\frac{s^3}{\sqrt{n}}\left(\frac{B^3A^3}{3(1-\rho)^3} + \frac{4B^2A^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{s}{\sqrt{n}}\right)_+^{-1}\right)
\end{aligned}$$

holds for some constants B and C. This is the situation of Theorem 1.9 except that in this case $d(n)$ is defined by

$$\frac{1}{\sqrt{n}}\frac{BCA^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB}-\frac{s}{\sqrt{n}}\right)_+^{-1} + \frac{1}{\sqrt{n}}\left(\frac{B^3A^3}{3(1-\rho)^3}+\frac{4B^2A^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB}-\frac{s}{\sqrt{n}}\right)_+^{-1}\right).$$

This expression depends on s. We apply the adapted Theorem 1.9 for $s_n=\frac{a_n^2}{n}$, $t_n:=\frac{a_n}{\sqrt{n}}$ and $s:=\lambda\frac{a_n}{\sqrt{n}}$ as before.

Because of $\frac{s}{\sqrt{n}}=\lambda\frac{a_n}{n}\overset{n\to\infty}{\longrightarrow}0$, the assumptions of Theorem 1.9 are satisfied:

1.
$$\begin{aligned}\frac{s_n^2}{t_n^3}d(n) &= \frac{a_n}{n}\frac{BCA^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB}-\frac{s}{\sqrt{n}}\right)_+^{-1}\\ &\quad+\frac{a_n}{n}\left(\frac{B^3A^3}{3(1-\rho)^3}+\frac{4B^2A^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB}-\frac{s}{\sqrt{n}}\right)_+^{-1}\right)\\ &\overset{n\to\infty}{\longrightarrow}0\,.\end{aligned}$$

2. $\frac{s_n}{t_n^2}\mathbb{V}f(X)=\mathbb{V}\big(\sqrt{n}U_n(h)(X)\big)<\infty$ as assumed.

Therefore we can use the Gärtner-Ellis theorem to prove the moderate deviation principle for $(\frac{n}{a_n}U_n(h)(X))_n$. □

COROLLARY **3.5.** *Let $(X_k)_{k\in\mathbb{N}}$ be a strictly stationary, aperiodic and irreducible Markov chain with finite state space and $U_n(h)(X)$ be a non-degenerate U-statistic based on a bounded kernel h of degree two. Then $(\frac{n}{a_n}U_n(h)(X))_n$ satisfies the MDP with speed and rate function as in Theorem 3.4.*

Proof. The Markov chain is strong mixing and the absolute regularity coefficient $\beta(n)$ converges to 0 at least exponentially fast as n tends to infinity, see [Bra05], Theorem 3.7(c). Hence the equations (3.19) and (3.20) are satisfied and Theorem 3.4 can be applied. The limit of the variance of $\sqrt{n}U_n(h)$ is bounded, see [Lee90], 2.4.2 Theorem 1, which proves the MDP for this example. □

For Doeblin recurrent and aperiodic Markov chains the MDP for additive functionals of a Markov process is proved in [Wu95]. In fact Wu proves the MDP under the condition that 1 is an isolated and simple eigenvalue of the transition probability kernel satisfying that it is the only eigenvalue with modulus 1. For a continuous spectrum of the transition probability kernel Delyon, Juditsky and Lipster present in [DJL06] a method for objects of the form

$$\frac{1}{n^\alpha}\sum_{i=1}^{n}H(X_{i-1}),\ \frac{1}{2}<\alpha<1, n\geq 1,$$

where $(X_i)_{i\geq 0}$ is a homogeneous ergodic Markov chain and the vector-valued function H satisfies a Lipschitz continuity. To the best of our knowledge, we proved the first MDP for a U-statistic with Markovian entries.

4 Proof of Theorem 1.1 and 1.3

LEMMA **4.1.** *The standardized subgraph count statistic Z satisfies the inequalities*

$$\Delta_i Z \leq \frac{1}{\sqrt{\binom{n}{2}p(1-p)}p^{k-1}} \tag{4.21}$$

$$\sum_{i=1}^{\binom{n}{2}} \sum_{j=1}^{i-1} \Delta_i \Delta_j Z \leq \frac{1}{c_{n,p}} \binom{n}{2} \binom{n-2}{l-2} (l-2)^2 (l-2)! \tag{4.22}$$

Proof of Lemma 4.1. As the first step we will find an upper bound for

$$\Delta_i Z = Z - \frac{1}{c_{n,p}} \sum_{1 \leq \kappa_1 < \cdots < \kappa_k \leq \binom{n}{2}} 1_{\{(e_{\kappa_1}, \ldots, e_{\kappa_k}) \sim G\}} \left(\prod_{j=1}^{k} X_{i,\kappa_j} - p^k \right),$$

where $(X_{i,1}, X_{i,2}, \ldots, X_{i,\binom{n}{2}}) = (X_1, \ldots, X_{i-1}, Y_i, X_{i+1}, \ldots, X_{\binom{n}{2}})$ and Y_i is an independent copy of X_i, $i \in \{1, \ldots, \binom{n}{2}\}$. The difference consists only of those summands which contain the random variable X_i or Y_i. The number of subgraphs isomorphic to G and containing a fixed edge, is given by

$$\binom{n-2}{l-2} \frac{2k}{a} (l-2)!,$$

see [NW88], p.307. Therefore we can estimate

$$\left| \Delta_i Z \right| \leq \frac{1}{\sqrt{\binom{n}{2} p(1-p)} p^{k-1}}. \tag{4.23}$$

For the second step we have to bound the partial difference of order two of the subgraph count statistic.

$$\begin{aligned}
& \Delta_i \Delta_j Z \\
= \; & \frac{1}{c_{n,p}} \sum_{\substack{1 \leq \kappa_1 < \cdots < \kappa_{k-2} \\ \kappa_1, \ldots, \kappa_{k-2} \neq i,j}} 1_{\{(e_i, e_j, e_{\kappa_1}, \ldots, e_{\kappa_{k-2}}) \sim G\}} \prod_{m=1}^{k-2} X_{\kappa_m} X_j (X_i - Y_i) \\
& - \frac{1}{c_{n,p}} \sum_{\substack{1 \leq \kappa_1 < \cdots < \kappa_{k-2} \\ \kappa_1, \ldots, \kappa_{k-2} \neq i,j}} 1_{\{(e_i, e_j, e_{\kappa_1}, \ldots, e_{\kappa_{k-2}}) \sim G\}} \prod_{m=1}^{k-2} X_{\kappa_m} Y_j (X_i - Y_i) \\
= \; & \frac{1}{c_{n,p}} \sum_{\substack{1 \leq \kappa_1 < \cdots < \kappa_{k-2} \\ \kappa_1, \ldots, \kappa_{k-2} \neq i,j}} 1_{\{(e_i, e_j, e_{\kappa_1}, \ldots, e_{\kappa_{k-2}}) \sim G\}} \prod_{m=1}^{k-2} X_{\kappa_m} (X_j - Y_j)(X_i - Y_i)
\end{aligned}$$

Instead of directly bounding the random variables we first care on cancellations due to the indicator function. We use the information about the fixed graph G. To do this we should distinguish the case, whether e_i and e_j have a common vertex.

- e_i and e_j have a common vertex:
  Because G contains l vertices, we have $\binom{n-3}{l-3}$ possibilities to fix all vertices of the subgraph isomorph to G and including the edges e_i and e_j. The order of the vertices is important and so we have to take the factor $2(l-2)!$ into account.

- e_i and e_j have four different vertices:
  Four fixed vertices allow us to choose only $\binom{n-4}{l-4}$ more. The order of the vertices and the relative position of e_i and e_j are relevant. So as before the factor is given by $2(l-2)!$.

Bounding the random variables X_i, Y_i, $i \in \{1, \ldots, \binom{n}{2}\}$ by 1, we achieve the following estimation:

$$\sum_{i=1}^{\binom{n}{2}} \sum_{j=1}^{i-1} \Delta_i \Delta_j Z \tag{4.24}$$

$$\leq \frac{\binom{n}{2}}{c_{n,p}} \left(4(n-2) \binom{n-3}{l-3} + (n-2)(n-3) \binom{n-4}{l-4} \right) (l-2)! \tag{4.25}$$

$$= \frac{1}{c_{n,p}} \binom{n}{2} \binom{n-2}{l-2} (l-2)^2 (l-2)! \, . \tag{4.26}$$

To bound $\sum_{j=1}^{i-1} \Delta_i \Delta_j Z$ for i fixed in (4.24) one has to observe that there are at most $2(n-2)$ indices $j < i$, such that e_i and e_j have a common vertex, and $\frac{1}{2}(n-2)(n-3) = \binom{n}{2} - 2(n-2) - 1$ indices j, such that e_i and e_j have no common vertex. This proves the inequality (4.25) and hence (4.26) follows. □

To apply Theorem 1.9 we choose $s_n = \frac{\left(\frac{2k}{a}(l-2)!\right)^2 \beta_n^2}{c_{n,p}^2}$ and $t_n = \frac{\beta_n}{c_{n,p}}$.

$$\frac{s_n}{t_n^2} \mathbb{V}Z = \left(\frac{2k}{a}(l-2)! \right)^2 \mathbb{V}Z \overset{n\to\infty}{\longrightarrow} \left(\frac{2k}{a}(l-2)! \right)^2 ,$$

because $\lim_{n\to\infty} \mathbb{V}Z = 1$, see [NW88]. We need Lemma 4.1 to bound $d(n)$:

$$d(n) \overset{(4.21)}{\leq} \frac{1}{\binom{n}{2} p^{2k-1}(1-p)} \sum_{i=1}^{\binom{n}{2}} \left(\frac{1}{3\sqrt{\binom{n}{2}} p^{k-1/2}(1-p)^{1/2}} + \sum_{j=1}^{i-1} \Delta_i \Delta_j Z \right)$$

$$\overset{(4.22)}{\leq} \frac{1}{\binom{n}{2} p^{2k-1}(1-p)} \left(\frac{\sqrt{\binom{n}{2}}}{3p^{k-1/2}(1-p)^{1/2}} + \frac{\binom{n}{2}\binom{n-2}{l-2}(l-2)^2(l-2)!}{c_{n,p}} \right)$$

$$= \frac{1}{\sqrt{\binom{n}{2}}} \frac{1}{p^{3(k-1/2)}(1-p)^{3/2}} \left(\frac{1}{3} + \frac{(l-2)^2 a}{2k} \right) \tag{4.27}$$

And condition 2 of Theorem 1.9 follows from

$$\frac{s_n^2}{t_n^3} d(n) \leq \beta_n \frac{1}{\binom{n-2}{l-2}\binom{n}{2}} \frac{1}{p^{4k-2}(1-p)^2} \left(\frac{1}{3} + \frac{(l-2)^2 a}{2k} \right) \left(\frac{2k}{a}(l-2)! \right)^3$$

$$\overset{n\to\infty}{\longrightarrow} 0 \quad \text{, if} \quad \beta_n \ll n^l p^{4k-2}(1-p)^2 \text{ as assumed.}$$

$\frac{s_n^2}{t_n^3} d(n)$ is positive and therefore the limit of n to infinity is zero, too. With Theorem 1.9 we proved Theorem 1.1.

REMARK 4.2. The estimation by Catoni, see Theorem 2.1, and Lemma 4.1 allow us to give an alternative proof for the central limit theorem of the subgraph count statistic Z, if $np^{3(k-\frac{1}{2})} \overset{n\to\infty}{\longrightarrow} \infty$ and $n^2(1-p)^{3/2} \overset{n\to\infty}{\longrightarrow} \infty$. On these conditions it follows, that $d(n) \overset{n\to\infty}{\longrightarrow} 0$, and it is easy to calculate the following limits:

$$\lim_{n\to\infty} \mathbb{E}e^{\lambda Z} = e^{\frac{\lambda^2}{2}\lim_{n\to\infty}\mathbb{V}Z} = e^{\frac{\lambda^2}{2}} < \infty \quad \text{for all } \lambda > 0$$
$$\text{and additionally} \quad \lim_{\lambda\nearrow 0}\lim_{n\to\infty} \mathbb{E}e^{\lambda Z} = 1\,.$$

Hence the central limit theorem results from the continuity theorem. Both conditions are stronger than the one in [NW88].

Proof of Theorem 1.3. We apply Theorem 2.1 and the Chebychev inequality to get

$$P(Z \geq \varepsilon) \leq \exp\left(-s\varepsilon + \frac{s^2}{2}\mathbb{V}Z + s^3 d(n)\right)$$

for all $s > 0$ and all $\varepsilon > 0$. Choosing $s = \frac{\varepsilon}{\mathbb{V}Z + \frac{2d(n)\varepsilon}{\mathbb{V}Z}}$ implies

$$P(Z \geq \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2\big(\mathbb{V}Z + \frac{2d(n)\varepsilon}{\mathbb{V}Z}\big)}\right). \tag{4.28}$$

Applying Theorem 2.1 to $-Z$ gives

$$P(Z \leq -\varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2\big(\mathbb{V}Z + \frac{2d(n)\varepsilon}{\mathbb{V}Z}\big)}\right).$$

Now we consider an upper bound for the upper tail $P(W - \mathbb{E}W \geq \varepsilon\,\mathbb{E}W) = P\big(Z \geq \frac{\varepsilon\,\mathbb{E}W}{c_{n,p}}\big)$. Using $\mathbb{V}Z = c_{n,p}^{-2}\mathbb{V}W$, inequality (4.28) leads to

$$P(W - \mathbb{E}W \geq \varepsilon\,\mathbb{E}W) \leq \exp\left(-\frac{\varepsilon^2(\mathbb{E}W)^2}{2\mathbb{V}W + 4\varepsilon d(n)\mathbb{E}W c_{n,p}^3(\mathbb{V}W)^{-1}}\right).$$

Indeed, this concentration inequality holds for $f(X) - \mathbb{E}f(X)$ in Theorem 1.9 with $d(n)$ given as in (1.12). We restrict our calculations to the subgraph-counting statistic W. We will use the following bounds for $\mathbb{E}W$, $\mathbb{V}W$ and $c_{n,p}$: there are constants, depending only on l and k, such that

$$const.\, n^l\, p^k \leq \mathbb{E}W \leq const.\, n^l\, p^k,$$
$$const.\, n^{2l-2}p^{2k-1}(1-p) \leq \mathbb{V}W \leq const.\, n^{2l-2}p^{2k-1}(1-p)$$

(see [Ruc88, 2nd section]), and

$$const.\, n^{l-1}p^{k-1/2}(1-p)^{1/2} \leq c_{n,p} \leq const.\, n^{l-1}p^{k-1/2}(1-p)^{1/2}.$$

Using the upper bound (4.27) for $d(n)$, we obtain

$$P(W - \mathbb{E}W \geq \varepsilon\,\mathbb{E}W) \leq \exp\left(-\frac{const.\,\varepsilon^2 n^{2l}p^{2k}}{n^{2l-2}p^{2k-1}(1-p) + const.\,\varepsilon n^{2l-2}p^{-k+1}(1-p)^{-1}}\right),$$

which proves Theorem 1.3. □

References

[BAG97] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997. MR1465640

[Bra05] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144 (electronic), 2005. Update of, and a supplement to, the 1986 original. MR2178042

[Cat03] O. Catoni. Laplace transform estimates and deviation inequalities. *Ann. Inst. H. Poincaré, Probabilités et Statistiques*, 39(1):1–26, 2003. MR1959840

[CD09] S. Chatterjee and P.S. Dey. Applications of Stein's method for concentration inequalities. *preprint, see arXiv:0906.1034v1*, 2009.

[DGZ03] A. Dembo, A. Guionnet, and O. Zeitouni. Moderate deviations for the spectral measure of certain random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(6):1013–1042, 2003. MR2010395

[DJL06] B. Delyon, A. Juditsky, and R. Liptser. Moderate deviation principle for ergodic Markov chain. Lipschitz summands. In *From stochastic calculus to mathematical finance*, pages 189–209. Springer, Berlin, 2006. MR2233540

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 1998. MR1619036

[Eic98] P. Eichelsbacher. Moderate and large deviations for U-processes. *Stochastic Process. Appl.*, 74(2):273–296, 1998. MR1624408

[Eic01] P. Eichelsbacher. Moderate deviations for functional U-processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(2):245–273, 2001. MR1819125

[ES03] P. Eichelsbacher and U. Schmock. Rank-dependent moderate deviations of u-empirical measures in strong topologies. *Probab. Theory Relat. Fields*, 126:61–90, 2003. MR1981633

[Jan90] S. Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990. MR1138428

[JŁR90] S. Janson, T. Łuczak, and T. Ruciński. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random graphs '87 (Poznań, 1987)*, pages 73–87. Wiley, Chichester, 1990. MR1094125

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. MR1782847

[JOR04] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004. MR2085711

[JR02] S. Janson and A. Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002. Probabilistic methods in combinatorial optimization. MR1900611

[JR04] S. Janson and A. Ruciński. The deletion method for upper tail estimates. *Combinatorica*, 24(4):615–640, 2004. MR2096818

[KV04] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004. MR2035874

[Lee90] A. J. Lee. *U–statistics Theory and Practice*, volume 110 of *Statistics: Textbooks and Monographs*. M. Dekker, New York, 1990. MR1075417

[NW88] K. Nowicki and J.C. Wierman. Subgraph counts in random graphs using incomplete *U*-statistics methods. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 299–310, 1988. MR0975550

[Ruc88] A. Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78(1):1–10, 1988. MR0940863

[Vu01] V. H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.*, 10(1):79–94, 2001. MR1827810

[Wu95] L. M. Wu. Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, 23(1):420–445, 1995. MR1330777