3 Applications I

This chapter consists of applications of the theory of large deviations as far as we have studied them in chapter 1 and 2. The first section deals with finite state irreducible Markov chains. In particular we will study the large deviations for the pair empirical measure.
The second section will introduce the Gibbs conditioning principle which is of fundamental importance in statistical mechanics. Shannon's source coding theorem is proved in Section 3.3 by combining the large deviations lower bound of the Gärtner - Ellis theorem. We conclude with samplepath large deviations for random walks and if time permits - for Brownian motions as well.

3.1 Large deviations for finite state Markov chains

We let $(X_n)_{n\geq 1}$ be a Markov chain taking values in a finite alphabet E, where without loss of generality E is identified with the set $\{1, .., N\}$, and $|E| = N$. Let $\Pi = (\pi_{i,j})_{i,j=1,..,N}$ be a stochastic matrix and denote by P_σ^Π the Markov probability measure associated with Π and with the initial state $\sigma \in E$, i.e.,

$$P_\sigma^\Pi(X_1 = x_1, \dots, X_n = x_n) = \pi(\sigma, x_1) \prod_{i=1}^{n-1} \pi(x_i, x_{i+1}) \; ; \; x_1, \dots, x_n \in E.$$

We let B^m denote the m-th power of the matrix. A matrix B with nonnegative entries is called irreducible, if for any pair of indices i,j there exists an $m = m(i,j)$ such that $B^m(i,j) > 0$ $\quad (B^m = (B^m(i,j))_{i,j=1,\dots,N})$.

We recall and state without proof the Perron-Frobenius theorem.

Theorem 11: (Perron-Frobenius)

Let $B = (B(i,j))_{i,j=1,\dots,N}$ irreducible matrix. Then B ~~possesses~~ possesses an eigenvalue ϱ such that:

(a) $\varrho > 0$ is real.

(b) For any eigenvalue λ of B, $|\lambda| \leq \varrho$

(c) For every $i \in E$ and every $\phi \in \mathbb{R}^{|E|}$ such that $\phi_j > 0$ for all $j \in E$,

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{j=1}^{|E|} B^n(i,j)\phi_j\right) = \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{j=1}^{|E|} \phi_j B^n(j,i)\right) = \log \varrho .$$

We first study large deviations of the empirical means

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} Y_k \quad , \text{where } Y_k = f(X_k) \text{ and}$$

$f: E \to \mathbb{R}$ is a given deterministic function.

The Gärtner-Ellis theorem hints that the rate function may still be expressed in terms of a Fenchel-Legendre transform, even when the random variables X_k obey a Markov dependence. Associate with every $\lambda \in \mathbb{R}$ a nonnegative matrix Π_λ, whose elements are

$$\Pi_\lambda(i,j) = \Pi(i,j)\, e^{\lambda f(j)} \quad , i,j \in E .$$

Clearly, Π_λ is irreducible as soon as Π is.
For each $\lambda \in \mathbb{R}$, let $\rho(\Pi_\lambda)$ denote the Perron-Frobenius's eigenvalue of the matrix Π_λ.

<u>Theorem 12:</u> Let $(X_n)_{n \geq 1}$ be a finite state Markov chain possessing an irreducible transition matrix Π.
For every $z \in \mathbb{R}$, define

$$I(z) := \sup_{\lambda \in \mathbb{R}} \left\{ \lambda z - \log \rho(\Pi_\lambda) \right\} .$$

Then the empirical mean Z_n satisfies the LDP with the convex, good rate function I.

<u>Proof:</u> In view of Theorem 10 (Gärtner-Ellis) respectively its finite dimensional version (i.e., Theorem 2.3.6 in the book by Dembo/Zeitouni), it suffices to check that the limit

$$\Lambda(\lambda) = \lim_{n \to \infty} \frac{1}{n} \Lambda_n(n\lambda) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}_\sigma^\Pi \left[e^{n \lambda Z_n} \right]$$

exists for every $\lambda \in \mathbb{R}$, that Λ is finite and differentiable everywhere in $\mathbb{R}$, and that $\Lambda(\lambda) = \log \rho(\Pi_\lambda)$.

$$\Lambda_n(n\lambda) = \log \mathbb{E}_\sigma^{\Pi}\left[e^{\lambda \sum_{k=1}^{N} Y_k} \right]$$

$$= \log \sum_{x_1,\dots,x_n \in E} P_\sigma^{\Pi}(Y_1 = x_1, \dots, Y_n = x_n) \prod_{k=1}^{n} e^{\lambda f(x_k)}$$

$$= \log \sum_{x_1,\dots,x_n \in E} \Pi(\sigma, x_1)\, e^{\lambda f(x_1)} \cdots \Pi(x_{n-1}, x_n)\, e^{\lambda f(x_n)}$$

$$= \log \sum_{x_n \in E} (\Pi_\lambda)^n (\sigma, x_n) .$$

We use (c) of Theorem 11 with $\phi = (1, \dots, 1)$ to get

$$\Lambda(\lambda) = \lim_{n \to \infty} \frac{1}{n} \Lambda_n(n\lambda) = \log \varrho(\Pi_\lambda) .$$

Moreover, since $|E|$ is finite, being an isolated root of the characteristic equation for the matrix Π_λ, is positive, finite and differentiable with respect to λ. ■

Recall the definition of the empirical measure L_n^X.

$$L_n^X(i) = \frac{1}{n} \sum_{k=1}^{n} 1\!\!1_{\{i\}}(X_k) \qquad , i = 1, \dots, |E| .$$

Here, L_n^X denotes the vector of frequencies in which the Markov chain visits the different states.

If Π is irreducible and μ the stationary distribution of the Markov chain (i.e., the unique left eigenvector of Π whose components sum to 1; the ergodic theorem implies that $L_n^X \to \mu$ in probability as $n \to \infty$, at least when Π is aperiodic and the initial state X_0 is sampled with μ.

Hence, the sequence $(L_n^X)_{n\geq 1}$ is a good candidate for an LDP in $\mathcal{M}_1(E)$.

We use now Theorem 12 with a function $f: E \to \mathbb{R}^{|E|}$ (this is a slight extension where we get scalar products $\langle \cdot, \cdot \rangle$ instead of products in our formulae), defined by

$f(y) = (1\!\!1_{\{1\}}(y), \ldots, 1\!\!1_{\{|E|\}}(y)) \in \mathbb{R}^{|E|}$. Therefore, by Theorem 12, the LDP holds for $(L_n^X)_{n\geq 1}$ with the rate function

$$I(q) = \sup_{\lambda \in \mathbb{R}^{|E|}} \left\{ \langle \lambda, q \rangle - \log \rho(\Pi_\lambda) \right\},$$

where $\Pi_\lambda(i,j) = \pi(i,j) e^{\lambda_j}$.

We shall give the following alternative characterisation of $I(q)$.

<u>**Proposition 13**</u>: $I(q) = J(q) := \begin{cases} \sup_{u \gg 0} \sum^{|E|} q_j \log \frac{u_j}{(u\Pi)_j}, & q \in \mathcal{M}_1(E) \\ \infty, & q \notin \mathcal{M}_1(E). \end{cases}$

Notation: $u \in \mathbb{R}^{|E|}$, $u \gg 0$ means $u_i > 0 \ \forall i \in E$.

Proof.: Lower bound of Theorem 12 yields for the open set $\mathcal{M}_1(E)^c$

$$-\infty = \liminf_{n\to\infty} 1/n \log P_\sigma^\pi (L_n^X \in \mathcal{M}_1(E)^c) \geq -\inf_{q \notin \mathcal{M}_1(E)} I(q) .$$

Consequently, $I(q) = \infty$ for every $q \notin \mathcal{M}_1(E)$.
Pick $q \in \mathcal{M}_1(E)$ and a strictly positive vector $u \gg 0, u \in \mathbb{R}^{|E|}$, and for $j = 1, \dots, |E|$, set $\lambda_j = \log(u_j / (u\pi)_j)$.
Observe that $u\pi_\lambda = u$, hence $u\pi_\lambda^n = u$ and thus $\rho(\pi_\lambda) = 1$ by part (c) of Theorem 11 (with $\phi_i = u_i > 0$).
Therefore, by definition,

$$I(q) \geq \sum_{j=1}^{|E|} q_j \log \frac{u_j}{(u\pi)_j} ,$$

henceforth $I(q) \geq J(q)$.
Pick $\lambda \in \mathbb{R}^{|E|}$ and let $u^* \gg 0$ be a left eigenvector to the eigenvalue $\rho(\pi_\lambda)$. Then, $u^*\pi_\lambda = \rho(M_\lambda) u^*$,
and

$$\langle \lambda, q \rangle + \sum_{j=1}^{|E|} q_j \log \frac{(u^*\pi)_j}{u_j^*} = \sum_{j=1}^{|E|} q_j \log \frac{(u^*\pi_\lambda)_j}{u_j^*}$$

$$= \sum_{j=1}^{|E|} q_j \log \rho(\pi_\lambda) = \log \rho(\pi_\lambda) . \text{ Therefore,}$$

$$\langle \lambda, q \rangle - \log \rho(\pi_\lambda) \leq J(q) . \qquad \blacksquare$$

<u>Remarks:</u> (1) If $(X_n)_{n\geq 1}$ are i.i.d., then the rows of Π are identical, in which case $J(q)$ is the relative entropy $H(q \mid \Pi(1,\cdot))$ (i.e., Sanov's theorem).

(2) $\Pi(i,j) = \mu(j)$, where $\mu \in \mathcal{M}_1(E)$.
Then $J(\cdot) = H(\cdot \mid \mu)$, and I is the Fenchel-Legendre transform of $\log\left(\sum_{j\in E} e^{x_j}\mu(j)\right)$.

The rate function governing the LDP for the empirical measure of a Markov chain is still in the form of an optimisation problem, however, the elegant interpretation in terms of relative entropy has disappeared.
Is there a somewhat different random variable, from which the LDP for $(L_n^X)_{n\geq 1}$ may be recovered, an LDP may be obtained with a rate function that is an appropriate relative entropy.
Consider the space $E^2 = E\times E$ and note that by considering the pairs formed by $X_1, \ldots, X_n$, i.e., the sequence $X_0X_1, X_1X_2, X_2X_3, \ldots, X_{n-1}, X_n$ where $X_0 = \sigma$, a Markov chain is recovered with state space E^2 and transition matrix $\Pi^{(2)}$

given as

$$\Pi^{(2)}((k,\ell),(i,j)) = 1\!\!1_{\{\ell\}}(i)\ \Pi(i,j).$$

We shall assume that Π is strictly positive (i.e., $\Pi(i,j)>0$). Then $\Pi^{(2)}$ is an irreducible transition matrix, and the results of Theorem 12 and Proposition 13 may be applied to yield the rate function $I_2(q)$ associated with the large deviations of the pair empirical measures

$$L^X_{n,2}(y) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{y\}}(X_{i-1},X_i),\ y\in E^2.$$

$L^X_{n,2} \in \mathcal{M}_1(E^2)$. We shall characterise I_2 as an appropriate relative entropy.

$q^{(1)}(i) = \sum_{j\in E} q(i,j)$ and $q^{(2)}(i) = \sum_{j\in E} q(j,i)$ denotes the first and second marginal of $q\in\mathcal{M}_1(E^2)$ respectively.

Let $q_f(j|i) = \dfrac{q(i,j)}{q^{(1)}(i)}$ whenever $q^{(1)}(i)>0$.

$q\in\mathcal{M}_1(E^2)$ is shift-invariant if $q^{(1)} = q^{(2)}$.

Theorem 14: Assume that Π is strictly positive. Then for any $q \in \mathcal{M}_1(E^2)$,

$$I_2(q) = \begin{cases} \sum_{i \in E} q^{(1)}(i)\, H(q_f(\cdot \mid i) \mid \pi(i,\cdot)) &, \text{if } q^{(1)} = q^{(2)} \\ \infty &, \text{otherwise.} \end{cases}$$

Remarks:

(1) When Π is not strictly positive, the theorem still applies, with E^2 replaced by $\{(i,j) : \pi(i,j) > 0\}$, and a similar proof.

(2) $I_2(q)$ is an expectation of relative entropies.

(3) The representation of $I_2(q)$ is useful in characterisation of the spectal radius of nonnegative matrices.
First, one can show that for any irreducible, nonnegative matrix B,

$$\log \rho(B) = \sup_{\nu \in \mathcal{M}_1(E)} \{ -J_B(\nu) \},$$

where J_B is the function in Proposition 13 for the matrix B.
Secondly, the contraction principle applied to $q \longmapsto q^{(2)}$ give for any strictly positive stochastic matrix Π

$$J(\nu) = \inf_{\substack{q \in \mathcal{M}_1(E^2): \\ q^{(2)} = \nu}} I_2(q), \quad \nu \in \mathcal{M}_1(E).$$

Finally, we combine the two steps to deduce that

$$-\log \varrho(B) = \inf_{\nu \in \mathcal{M}_1(E)} \{J(\nu)\} = \inf_{\nu \in \mathcal{M}_1(E)} \inf_{\substack{q \in \mathcal{M}_1(E^2): \\ q^{(2)} = \nu}} I_2(q)$$

$$= \inf_{\substack{q \in \mathcal{M}_1(E^2): \\ q^{(1)} = q^{(2)}}} \sum_{(i,j)} q(i,j) \log \frac{q_f(j|i)}{b(i,j)} .$$

This is Varadhan's characterisation of the spectral radius of nonnegative irreducible matrices.

Proof of Theorem 14:

By Proposition 13 we have

$$I_2(q) = \sup_{u \gg 0} \sum_{i,j \in E} q(i,j) \log \underbrace{\frac{u(i,j)}{(u\Pi^{(2)})(i,j)}}_{\log \frac{u(i,j)}{\left(\sum_{k \in E} u(k,i)\right) \pi(i,j)}} .$$

Pick $q \in \mathcal{M}_1(E^2)$ such that $q^{(1)} \neq q^{(2)}$.

$q^{(1)}(j_0) < q^{(2)}(j_0)$ for some $j_0 \in E$. For u such that $u(\cdot, j) = 1$ when $j \neq j_0$ and $u(\cdot, j_0) = e^{\alpha}$, we get

$$\sum_{i,j \in E} q(i,j) \log \frac{u(i,j)}{\left(\sum_{k \in E} u(k,i)\right) \pi(i,j)} = \sum_{i,j \in E} q(i,j) \log \frac{u(1,j)}{|E|\, u(1,i)\, \pi(i,j)}$$

$$= -\sum_{i,j\in E} q(i,j)\log\left(|E|\,\pi(i,j)\right) \quad + \alpha\left(q_2(j_0)-q_1(j_0)\right).$$

Letting $\alpha\to\infty$, we find that $I_2(q)=\infty$.

If $q^{(1)}=q^{(2)}$, then for every $u\gg 0$,

$$\sum_{i,j\in E} q(i,j)\log\frac{\sum_{k\in E}u(k,i)\,q^{(2)}(j)}{\sum_{k\in E}u(k,j)\,q^{(1)}(i)} = 0.$$

Let $u(i|j)=u(i,j)\,/\sum_{k\in E}u(k,j)$ and $q_b(i|j):=\frac{q(i,j)}{q^{(2)}(j)}$.

By the last two displays we arrive at

$$I_2(q)-\sum_{i\in E} q^{(1)}(i)\,H\left(q_f(\cdot|i)\,|\,\pi(i,\cdot)\right)$$

$$=\sup_{u\gg 0}\sum_{i,j\in E} q(i,j)\log\frac{u(i,j)\,q^{(1)}(i)}{\left(\sum_{k\in E}u(k,i)\right)q(i,j)}$$

$$=\sup_{u\gg 0}\sum_{i,j\in E} q(i,j)\log\frac{u(i|j)}{q_b(i|j)} = \sup_{u\gg 0}\left\{-\sum_{j\in E} q^{(2)}(j)\,H\left(q_b(\cdot|j)\,|\,u(\cdot|j)\right)\right\}$$

Note that always $I_2(q)\le\sum q^{(1)}(i)\,H\left(q_f(\cdot|i)\,|\,\pi(i,\cdot)\right)$.

(choice $u=q$ yields equality). The proof is completed for q, which is not strictly positive, by considering a sequence $u_n\gg 0$ such that $u_n\to q$. ∎