1. Markov Chains.

Let us switch from independent to dependent random variables. Suppose $X_1, \cdots, X_n, \cdots$ is a Markov Chain on a finite state spacce F consisting of points $x \in F$. The Markov Chain will be assumed to have a stationary transition probability given by a stochastic matrix $\pi = \pi(x, y)$, as the probability for transition from x to y. We will assume that all the entries of π are positive, imposing thereby a strong irreducibility condition on the Markov Chain. Under these conditions there is a unique invariant or stationary distribution $p(x)$ satisfying

$$\sum_x p(x)\pi(x, y) = p(y).$$

Let us suppose that $V(x) : F \to R$ is a function defined on the state space with a mean value of $m = \sum_x V(x)p(x)$ with respect to the invariant distribution. By the ergodic theorem, for any starting point x,

$$\lim_{n\to\infty} P_x\big[|\frac{1}{n}\sum_j V(X_j) - m| \geq a\big] = 0$$

where $a > 0$ is arbitrary and P_x denotes, as is customary, the measure corresponding to the Markov Chain initialized to start from the point $x \in F$. We expect the rates to be exponential and the goal as it was in Cramer's theorem, is to calculate

$$\lim_{n\to\infty} \frac{1}{n} \log P_x\big[\frac{1}{n}\sum_j V(X_j) \geq a\big]$$

for $a > m$. First, let us remark that for any V

$$\lim_{n\to\infty} \frac{1}{n} \log E_x\Big[\exp[V(X_1) + V(X_2) \cdots + V(X_n)]\Big] = \log \sigma(V)$$

exists where $\sigma(V)$ is the eigenvalue of the matrix

$$\pi_V = \pi_V(x, y) = \pi(x, y)e^{V(x)}$$

with the largest modulus, which is positive. Such an eigenvalue exists by Frobenius theory and is characterized by the fact that the eigenvector has positive entries. To see this we only need to write down the formula

$$E_x\Big[\exp[V(X_1) + V(X_2) \cdots + V(X_n)]\Big] = \sum_y [\pi_V]^n(x, y)$$

that is easily proved by induction. The geometric rate of growth of any of the entries of the n-th power of a positive matrix is of course given by the Frobenius eigenvalue σ. Once we know that

$$\lim_{n\to\infty} \frac{1}{n} \log E_x\Big[\exp[V(X_1) + V(X_2) \cdots + V(X_n)]\Big] = \log \sigma(V)$$

we can get an upper bound by Tchebychev's inequality

$$\limsup_{n\to\infty}\frac{1}{n}\log P_x\Big[V(X_1)+V(X_2)\cdots+V(X_n)\ge a\Big]\le \log\sigma(V)-a$$

or repalcing V by λV with $\lambda>0$

$$\limsup_{n\to\infty}\frac{1}{n}\log P_x\Big[V(X_1)+V(X_2)\cdots+V(X_n)\ge a\Big]\le \log\sigma(\lambda V)-a\lambda$$

Since we can optimize over $\lambda\ge 0$ we can obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log P_x\Big[V(X_1)+V(X_2)\cdots+V(X_n)\ge a\Big]\le -h(a)=-\sup_{\lambda\ge 0}[\lambda a-\log\sigma(\lambda V)]$$

By Jensen's inequality

$$\begin{aligned}\sigma(\lambda V)&=\lim_{n\to\infty}\frac{1}{n}\log E^{P_x}\Big[\exp\big[\lambda[V(X_1)+V(X_2)\cdots+V(X_n)]\big]\Big]\\&\ge\lim_{n\to\infty}\frac{1}{n}E^{P_x}\Big[\lambda[V(X_1)+V(X_2)\cdots+V(X_n)]\Big]\\&=\lambda m\end{aligned}$$

Therefore

$$h(a)=\sup_{\lambda\ge 0}[\lambda a-\log\sigma(\lambda V)]=\sup_{\lambda\in R}[\lambda a-\log\sigma(\lambda V)]$$

For the lower bound we can again perform the trick of Cramér to change the measure from P_x to a Q_x such that under Q_x the event in question has probability nearly 1. The Radon Nikodym derivative of P_x with respect to Q_x will then control the large deviation lower bound. Our plan then is to replace π by $\bar\pi$ such that $\bar\pi(x,y)$ has an invariant measure $q(x)$ with $\sum_x V(x)q(x)=a$. If Q_x is the distribution of the chain with transition probability $\bar\pi$ then

$$P_x[E]=\int_E R_n dQ_x$$

where

$$R_n=\exp\big[\sum_j\log\frac{\pi(X_j,X_{j+1})}{\bar\pi(X_j,X_{j+1})}\big]$$

This provides us a lower bound for the probbaility

$$-\liminf_{n\to\infty}\frac{1}{n}\log P_x\Big[V(X_1)+V(X_2)\cdots+V(X_n)\ge a\Big]\le J(\bar\pi)$$

where

$$\begin{aligned}J(\bar\pi)&=\lim_{n\to\infty}\frac{1}{n}E^Q_x[-\log R_n]\\&=\sum_{x,y}\log\frac{\bar\pi(x,y)}{\pi(x,y)}\bar\pi(x,y)q(x)\end{aligned}$$

The last step comes from applying the ergodic theorem for the $\bar{\pi}$ chain with $q(\cdot)$ as the invariant distribution to the average

$$-\frac{1}{n}\log R_n = \frac{1}{n}\sum_{j=1}^{n}\log\frac{\bar{\pi}(X_j, X_{j+1})}{\pi(X_j, X_{j+1})}$$

Since any $\bar{\pi}$ with $\sum_x V(x)q(x) = a$ will do where $q(\cdot)$ is the corresponding invariant measure, the lower bound we have is quickly improved to

$$\lim_{n\to\infty}\frac{1}{n}\log P_x\left[V(X_1)+V(X_2)\cdots+V(X_n)\geq a\right] \geq -\inf_{\substack{\bar{\pi}\\ \sum V(x)q(x)=a}} J(\bar{\pi})$$

If we find the λ such that $\lambda a - \log\sigma(\lambda V) = h(a)$ then for such a choice of λ

$$a = \frac{\sigma'(\lambda V)}{\sigma(\lambda V)}$$

and we take $\pi(x,y)e^{\lambda V(y)}$ as our nonnegative matrix. We can find for the eigenvalue σ a column eigenvector f and a row eigenvector g. We define

$$\bar{\pi}(x,y) = \frac{1}{\sigma}\pi(x,y)e^{\lambda V(y)}\frac{f(y)}{f(x)}$$

One can check that $\bar{\pi}$ is a stochastic matrix with invariant measure $q(x) = g(x)f(x)$ (properly normalized). An elemenatary perturbation argument involving eigenvalues gives the relationship

$$a = \sum_x q(x)V(x)$$

and

$$\begin{aligned} J &= \sum_{x,y}[-\log\sigma + \lambda V(y) + \log f(y) - \log f(x)]\bar{\pi}(x,y)q(x) \\ &= \lambda a - \log\sigma \\ &= h(a) \end{aligned}$$

thereby matching the upper bound. We have therefore proved the following theorem.

Theorem 1.1. *For any Markov Chain with a transition probability matrix π with positive entries, the probability distribution of $\frac{1}{n}\sum_{j=1}^{n}V(X_j)$ satisfies an LDP with a rate function*

$$h(a) = \sup_{\lambda}[\lambda a - \log\sigma(\lambda V)]$$

There is an interesting way of looking at $\sigma(V)$. If $V(x) = \log\frac{u(x)}{(\pi u)(x)}$ then $f(x) = (\pi u)(x)$ is a column eigenfunction for π_V with eigenvalue $\sigma = 1$. Therefore

$$\log\sigma(\log\frac{u}{\pi u}) = 0$$

and because $\log \sigma(V+c) = \log \sigma(V) + c$ for any constant c, $\log \sigma(V)$ is the amount c by which V has to be shifted so that $V - c$ is in the class

$$\mathcal{M} = \{V : V = \log \frac{u}{\pi u}\}$$

We now turn our attention to the number of visits

$$L_x^n = \sum_{j=1}^{n} \chi_x(X_j)$$

to the state x or

$$\ell_x^n = \frac{1}{n} L_x^n$$

the proportion of time spent in the state x. If we view $\{\ell_x; x \in F\}$ as a point in the space $\mathcal{P}$ of probability measures on F, then we get a distribution ν_x^n for any starting point x and time n. The ergodic theorem asserts that ν_x^n converges weakly to the degenerate distribution at $p(x)$, the invariant measure for the chain. In fact we have actually proved the large deviation principle for ν_x^n.

Theorem 1.2. *The distributions ν_x^n of the empirical distributions satisfy an LDP with a rate function*

$$I(q) = \sup_{V \in \mathcal{M}} \sum_x q(x) V(x)$$

Proof. Upper Bound. Let $q \in \mathcal{P}$ be arbitrary. Suppose $V \in \mathcal{M}$. Then

$$E^{P_x}[\exp[\sum V(X_i)](\pi u)(X_{n+1})] = (\pi u)(x)$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log E_x \big[\exp[\sum_{j=1}^{n} V(X_j)] \big] \leq 0$$

which implies that

$$\limsup_{U \downarrow q} \limsup_{n\to\infty} \frac{1}{n} \log \nu_x^n[U] \leq -\sum_x q(x) V(x)$$

where U are neighborhoods of q that shrink to U. Since the space $\mathcal{P}$ is compact this provides immediately the upper bound in the LDP with the rate function $I(q)$. The lower bound is a matter of finding a $\bar{\pi}$ such that $q(x)$ is the invariant measure for $\bar{\pi}$ and $J(\bar{\pi}) = I(q)$. One can do the variational problem of minimizing $J(\bar{\pi})$ over those $\bar{\pi}$ that have q as an invariant measure. The minimum is attained at a point where

$$\bar{\pi}(x,y) = \pi(x,y) \frac{f(y)}{(\pi f)(x)}$$

with q as the invariant measure for $\bar{\pi}$. Since $q(x)$ is the the invariant distribution for $\bar{\pi}$ an easy calculation gives

$$\begin{aligned} J &= \sum_{x,y} [\log f(y) - \log(\pi u)(x)]\bar{\pi}(x,y)q(x) \\ &= \sum_{x} [\log f(x) - \log(\pi u)(x)]q(x) \\ &= \sum_{x} V(x)q(x) \text{ for some } V \in \mathcal{M} \\ &\leq I(q) \end{aligned}$$

Since we already have the upper bound with $I(\cdot)$ we are done. □

Remark 1.3. It is important to note that in the special case of IID random variables with a finite set of possible values we can take $\pi(x,y) = \pi(y)$ and for any $V(x)$ the matrix $\pi(y)e^{V(y)}$ is of rank one with the nonzero eigenvalue being equal to $\sigma(V) = \sum_x e^{V(x)}\pi(x)$ and the analysis reduces to the one of Cramer. In particular

$$I(q) = \sum_x q(x) \log \frac{q(x)}{\pi(x)}$$

is the relative entropy with respect to $\pi(\cdot)$.

Remark 1.4. A similar analysis can be done for the continuous time Markov chains as well. Suppose we have a Markov chain, again on the finite state space F, with transition probabilities $\pi(t,x,y)$ given by $\pi(t) = \exp[tA]$ where the infinitesiaml generator or the rate matrix A satisfies $a(x,y) \geq 0$ for $x \neq y$ and $\sum_{y \in F} a(x,y) = 0$. For any function $V : F \to R$ the operator $(A + V)$ defined by

$$(A + V)f(x) = \sum_y a(x,y)f(y) + V(x)f(x)$$

has the property that the eigenvalue with the largest real part is real and if we make the strong irreducibility asssumption as before i.e. $a(x,y) > 0$ for $x \neq y$, then this eigenvalue is unique and is characterized by having row (or equivalently) column eigenfunctions that are positive. We denote this eigenvalue by $\theta(V)$. If $u > 0$ is a positive function on F, then $Au + (\frac{-Au}{u})u = Au - Au = 0$. Therefore for $V = \frac{-Au}{u}$, the vector u is a column eigenvector for the eigenvalue 0 and for such V, $\theta(V) = 0$. If we now denote by $[\mathcal{M} = V : V = \frac{-Au}{u}$ for some $u > 0\}$ then $\theta(V)$ is the unique constant such that $V - \theta \in \mathcal{M}$. We can now have the exact analogues of the descrete case and with the rate function

$$I(q) = \sup_{V \in \mathcal{M}} \sum_x q(x)V(x)$$

we have an LDP for the empirical measures

$$\ell_x(t) = \frac{1}{t} \int_0^t \chi_x(X(s))ds$$

that represent the proportion of time spent in the state $x \in F$.

Remark 1.5. In the special case when A is symmetric one can explicitly calculate

$$I(q) = -\sum_{x,y} a(x,y)\sqrt{q}(x)\sqrt{q}(y).$$

The large deviation theory gives the asymptotic relation

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t}\log E\big[\exp\big[\int_0^t V(X(s))ds\big]\big] &= \sup_{q(\cdot)\in\mathcal{P}} \Big[\sum_x q(x)V(x) - I(q)\Big] \\ &= \sup_{\substack{f:f\geq 0\\ \|f\|_2=1}} \sum_x V(x)[f(x)]^2 + \sum_{x,y} a(x,y)f(x)f(y) \\ &= \sup_{f:\|f\|_2=1} \sum_x V(x)[f(x)]^2 + \sum_{x,y} a(x,y)f(x)f(y)\end{aligned}$$

Notice that the Dirichlet form $\sum_{x,y} a(x,y)f(x)f(y)$ can be rewritten as $\frac{1}{2}\sum_{x,y} a(x,y)[f(x)-f(y)]^2$ and replacing f by $|f|$ lowers the Dirichlet form and so we have dropped the assumption that $f \geq 0$. This is the familiar variational formula for the largest eigenvalue of a symmetric matrix.