1. Empirical Processes.

Let us return to the situation of a Markov chain on a finite state space X, having $\pi(x,y)$ as the probability of transition from the state x to the state y. We saw before that

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

exists and is independent of x. More over the limit $\lambda(V)$ is given by a variational formula

$$\lambda(V)=\sup_{q\in\mathcal{P}}\big[\sum_x V(x)q(x)-I(q)\big]$$

where $q=\{q(x)\}$ is a probability distribution on X, $\mathcal{P}$ is the space of such probability distributions and $I(q)$ is the large deviation rate function for the distibution Q_n on $\mathcal{P}$, of the empirical distribution

$$p_n(x)=\frac{1}{n}\sum_{i=1}^{n}\chi_x(X_i).$$

We want to generalize this to study limits of the form

$$\lambda_2(V)=\lim_{n\to\infty}\frac{1}{n}\log E_x\big[\exp[V(X_1,X_2)+V(X_2,X_3)+\cdots+V(X_n,X_{n+1})]\big].$$

To this end we construct a Markov Process on a new state space $X^{(2)}=X\times X$ where the transitions from (x_1,x_2) to (y_1,y_2) are possible only if $x_2=y_1$ and the probability of the transition is $\pi(x_2,y_2)$. In other words we carry a memory of one previous state as part of the current state. Although the transition probabilities on this enlarged state space are not all positive, they all become positive after two iterations and this is sufficient to extend the previous analysis to this case. If $q_2(\cdot,\cdot)$ is a probability on $X^{(2)}$, the rate function $I_2(q_2)$ is finite if and only if $\sum_x q_2(x,a)=\sum_x q_2(a,x)$ for all a, i.e the two one dimensional marginals of q_2 are identical. For such a q_2 we can write $q_2(x,y)=\alpha(x)\beta(y|x)$ interms of a marginal and a conditional. Then a calculation shows that

$$I_2(q_2)=\sum_x\alpha(x)H(x)$$

where H is the relative entropy

$$H(x)=\sum_y\beta(y|x)\log\frac{\beta(y|x)}{\pi(x,y)}.$$

Now the rate function is more explicit than before and the connection is through the contraction formula

$$I(q)=\inf_{q_2\in\mathcal{P}_q}I_2(q_2)$$

where $\mathcal{P}_q$ are those probability distributions $X \times X$ that have both marginals equal to the given distibution q on X. This can be extended to consideration of limits of the form

$$\lambda_k(V) = \lim_{n\to\infty} \frac{1}{n} \log E_x\big[\exp[V(X_1, X_2, \cdots, X_k) + V(X_2, X_3, \cdots, X_{k+1}) + \cdots V(X_n, X_{n+1}, \cdots, X_{n+k-1})]\big]$$

by means of a rate functional $I_k(q_k)$ on probability distributions q_k on $X^{(k)} = X \times X \cdots \times X$ that is finite only for those q_k that have the same projection on the first $k-1$ components as the last $k-1$. The q_k is written as

$$q_k(x_1, x_2, \cdots, x_k) = \alpha(x_1, x_2, \cdots, x_{k-1})\beta(x_k|x_1, x_2, \cdots, x_{k-1})$$

and

$$I_k(q_k) = \sum_{x_1, x_2, \cdots, x_{k-1}} \alpha(x_1, x_2, \cdots, x_{k-1}) H(x_1, x_2, \cdots, x_{k-1})$$

where

$$H(x_1, x_2, \cdots, x_{k-1}) = \sum_{x_k} \beta(x_k|x_1, x_2, \cdots, x_{k-1}) \log \frac{\beta(x_k|x_1, x_2, \cdots, x_{k-1})}{\pi(x_{k-1}, x_k)}$$

is the relative entropy. The final step is to combine it all with a rate function $I(\mu)$ on the space $\mathcal{M}$ of stationary X-valued stochastic processes defined by

$$I(\mu) = E_\mu[H(\omega)]$$

where

$$H(\omega) = \sum_x \beta(x|\omega) \log \frac{\beta(x|\omega)}{\pi(X_0(\omega), x)}$$

is the relative entropy of the conditional distribution

$$\beta(x|\omega) = \mu[X_1 = x | X_0, X_{-1}, \cdots]$$

of the process at time 1 given the past upto time 0 with respect to the Markovian conditional distribution $\pi(X_0(\omega), x)$ where $X_0(\omega)$ is the 0-th coordinate that is part of the past. If we denote by Π_k the projection from $\Omega \to X^{(k)}$, as well as the induced map from measures on Ω to measures on $X^{(k)}$

$$I_k(q_k) = \inf_{\mu : \Pi_k \mu = q_k} I(\mu).$$

Remark: One of the applications of such a general formula is to the study of limits of the type

$$\lim_{n\to\infty} \frac{1}{n} \log E_x\Big[\exp \sum_{i,j=1}^{n} \big[\rho^{|i-j|} V(X_i, X_j)\big]\Big]$$

for a symmetric function V on $X \times X$.

Proof. Let us do the independent case. $\{X_i\}$ are i.i.d with a common distribution $\pi(dx)$ We begin by estimating

$$\frac{1}{n}\log E^P[\exp[\sum_{i=1}^{n} V(X_i,\ldots,X_{i+k-1}]]$$

for an any bounded continuous function $V(x_1,x_2,\ldots,x_k) : X^k \to R$ where k is any positive integer. Replace $n = nk$

$$\begin{aligned}\exp[\sum_{i=1}^{nk} V(X_i,\ldots,X_{i+l-1}] &= \exp[\sum_{j=1}^{k}\sum_{i=1}^{n} V(X_{(i-1)k+j},X_{(i-1)k+j+1},\ldots,X_{ik+j-1})] \\ &\le \frac{1}{k}\sum_{j=1}^{k}\exp[k\sum_{i=1}^{n} V(X_{(i-1)k+j},X_{(i-1)k+j+1},\ldots,X_{ik+j-1})]\end{aligned}$$

$$\begin{aligned}E^P[\exp[\sum_{i=1}^{nk} V(X_i,\ldots,X_{i+l-1}]] &\le E^P[\exp[k\sum_{i=1}^{n} V(X_{(i-1)k+j},X_{(i-1)k+j+1},\ldots,X_{ik+j-1})]] \\ &= \Big[E^P[\exp[kV(X_1,\ldots,X_k)]]\Big]^n\end{aligned}$$

$$\limsup_{n\to\infty}\frac{1}{n}\log E^P[\exp[\sum_{i=1}^{n} V(X_i,\ldots,X_{i+k-1})]] \le \frac{1}{k}\log E^P[\exp[kV(X_1,\ldots,X_k)]]$$

If $X_1,X_2,\ldots,X_n$ is a realization, we extend it periodically both sides to get a periodic random infinite sequence

$$\omega_n = (\ldots,X_1,X_2,\ldots,X_n,X_1,X_2,\ldots,X_n,X_1,\ldots,\ldots)$$

It defines a periodic orbit of period n and the orbital measure $R_n = \frac{1}{n}\sum_{j=0}^{n-1}\delta_{T^j\omega_n}$ defines a random stationary process that we call the empirical process. Its marginals are the empirical distributions of $\{X_i\}$, $\{X_i,X_{i+1}\},\ldots,\{X_i,X_{i+1},\ldots X_{i+k-1}\}$ etc except for the effect of periodization at the edges which is not significant if k is fixed and n is large. We now have bounds

$$\limsup_{N\to\infty}\frac{1}{n}E^P[\exp[n\int V(x_1,x_2,\ldots,x_k)dR_n]] \le \frac{1}{k}\log E^P[\exp[kV(X_1,\ldots,X_k)]]$$

It provides a large deviation upper bound with rate function

$$\begin{aligned} I_P(R) &= \sup_{k,V}\left[\int V(x_1,x_2,\ldots,x_k)dR - \frac{1}{k}\log E^P[\exp[kV(X_1,\ldots,X_k)]]\right] \\ &= \sup_{k,V}\left[\frac{1}{k}\int V(x_1,x_2,\ldots,x_k)dR - \frac{1}{k}\log E^P[\exp[V(X_1,\ldots,X_k)]]\right] \\ &= \sup_{k,V}\frac{1}{k}\left[\int V(x_1,x_2,\ldots,x_k)dR - \log E^P[\exp[V(X_1,\ldots,X_k)]]\right] \\ &= \sup_{k}\frac{1}{k}\left[\int V(x_1,x_2,\ldots,x_k)dR - \log E^P[\exp[V(X_1,\ldots,X_k)]]\right] \\ &= \sup_{k}\frac{1}{k}H(R_k|P_k) \\ &= \int H(R_\omega(dx);\alpha)dR \end{aligned}$$

where R_ω is the conditional distribution of X_1 given the past and α is the common distribution of, $\{X_i\}$ i.e. the one dimensional marginal of P. □

The last step follows from the following observation.

$$H(R_n|P_n) = E[H(R_1|\alpha)] + H(R_2,\omega)|\alpha) + H(R_n,\omega)|\alpha)]$$

where $R_{j,\omega}$ for $j \geq 2$ is the conditional distribution of x_j under R given $(x_1,\ldots,x_{j-1})$. $E[H(R_{n,\omega})]$ is nondecreasing and its limit is easily seen to be $E[H(R_\omega|\alpha)]$ by stationarity, martingale convergence theorem. Stationarity reduces the calculation to showing

$$\lim_{n\to\infty} E^R[H(R_{[-n,0],\omega}(dx_1)|\alpha)] = \int H(R_\omega(dx);\alpha)dR$$

The measures $R_{[-n,0],\omega}(dx_1)$ converge to $R_\omega(dx)$ by the martingale convergence theorem. The entropy is a convex lower semi continuous functional. Therefore almost surely

$$\liminf_{n\to\infty} H(R_{[-n,0],\omega}(dx_1)|\alpha) \geq H(R_\omega(dx)|\alpha)$$

and by Fatou's lemma

$$\liminf_{n\to\infty} E^R[H(R_{[-n,0],\omega}(dx_1)|\alpha)] \geq \int H(R_\omega(dx)|\alpha)dR$$

On the other hand by the properties of entropy

$$
\begin{aligned}
E^R[H(R_{[-n,0],\omega}(dx_1)|\alpha)] = &\sup_{V=V(x_{-n},\ldots,x_0,x_1)} \Big[E^R[V(x_{-n},\ldots,x_0,x_1) \\
&- \log E^R\big[\int \exp[V(x_{-n},\ldots,x_0,x_1)]\alpha(dx_1)\big]\Big] \\
\le &\sup_V \Big[E^R[V(\omega,x_1)] - \log E^R\big[\int \exp[V(\omega,x_1)]\alpha(dx_1)\big]\Big] \\
= &\int H(R_\omega(dx)|\alpha)dR
\end{aligned}
$$

Now we turn to the proof of the lower bound. We first note that entropy $I_P(R)$ is linear in R. The conditional probability $R_\omega(dx_1)$ can be chosen universally. The is a bit of soft, but correct reasoning. The ergodic measures live on mutually disjoint subsets. So define R_ω differently on different sets, but call it the same. It then works for convex combinations which is every thing. Therefore it is enough to prove the lower bound for R that is ergodic. Neighborhoods are defined by finite dimensional distributions. Let us say we want k dimensional distribution of the empirical process to be close to that of R. There is a $(k-1)$ Markov process R_k with k dimensional distribution matching that of R. $I_P(R_k) \le I_P(R)$. It has a conditional density $r_k(x_k|x_1,x_2,\ldots,x_{k-1})$ with respect to α and if we tilt by $\exp\sum_{j=0}^{n-1}\log r_k(x_{j+k}|x_1,x_2,\ldots,x_{j+k-1})$ we get the lower bound.