**Large Deviations. Averaged case.**

To carry out large deviations, we cannot ignore events that have exponentially small probaility. In particular the process can behave like a recurrent one. So the space $\overline{\mathcal{W}}$ has to be enlarged to include recurrent paths. This makes $q(z|w)$ unstable. It is continuous, even when all paths are transient, only if the walks $w^{(j)}$ converge to a limit $w$ with the total number of visits $k_x(w^{(j)})$ of $w^{(j)}$ to $x$ converges to the number of visits $k_x(w)$ of $w$ to $x$ for every $x$. It is possible in the limit for $k_x(w)$ to be smaller. This requires a careful comapctification of $\overline{\mathcal{W}}$ to get $\mathbf{W}$.

We denote by $\mathcal{I}$ the set of probability measures on $\overline{\mathcal{W}}$ that are invariant under $T^*$. $T^*$ corresponds to erasing the last jump and shifting to end at the origin. $T^*T_z = Id$. These correspond to transient processes with stationary increments. We get a shift invariant measure on $\mathbf{z} = \{z_j\}$, initially for $j \leq 0$, which is easily extended to $-\infty < j < \infty$ to be shift invariant in both directions. The ergodic measures are the extremals in $\mathcal{I}$ and these are denoted by $\mathcal{E}$. If $\mu \in \mathcal{I}$, the conditional probability $q_\mu(z|w)$ of the next jump being equal to $z$ given the past history is well defined. Standard ergodic theory even yields a version $\hat{q}$, of this conditional probability that is universal, i.e. independent of $\mu$. But this is 'cheating' because the ergodic or extremal $\mu$'s are supported on mutually disjoint sets and one defines $\hat{q}(z|w)$ to equal $q_\mu(z|w)$ on the support of $\mu$. There is a natural large deviation rate functional in this context given by

$$I(\mu) = \int_{\overline{\mathcal{W}}} [\sum_z \hat{q}(z|w) \log \frac{\hat{q}(z|w)}{q(z|w)}] \mu(dw)$$

Note that one consequence of the universality of $\hat{q}$ is the affine linearity of $I(\cdot)$ on $\mathcal{I}$. Of course, each $\mu \in \mathcal{I}$ being a process with stationary increments has a mean 'drift' $m(\mu) = E^\mu[z_0]$. It is natural to define by contraction, especially for $\xi \neq 0$, the rate function

$$H(\xi) = \inf_{\substack{\mu \in \mathcal{E} \\ m(\mu) = \xi}} I(\mu)$$

We can now state the main theorem regarding the 'averaged' measure $\bar{Q}$.

**Theorem 1.** The function $H(\xi)$ defined above for $\xi \neq 0$ extends to $R^d$ as a convex function and is the rate function for large deviations of $\frac{X_n}{n}$ under 'averaged' measure $\bar{Q}$.

$$\limsup_{n\to\infty} \frac{1}{n} \log \bar{Q}\big[\frac{X_n}{n} \in C\big] \leq -\inf_{\xi \in C} H(\xi) \quad \text{for closed} \quad C \subset R^d$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \bar{Q}\big[\frac{X_n}{n} \in G\big] \leq -\inf_{\xi \in G} H(\xi) \quad \text{for open} \quad G \subset R^d$$

While the lower bound is routine, the upper bound requires some careful handling due to the lack of sufficient compactness and continuity of $q(z|w)$.

**Compactification.** There are various sets of walks. $\mathcal{W} = \cup_n \mathcal{W}_n$ is the space of all walks of finite length. We have $\overline{\mathcal{W}}$ obtained by adding all transient infinite walks to the collecetion. For large deviations we need to consider recurrent walks as well. We denote by $\mathbf{W}$ the space obtained by adding all infinite walks to $\mathcal{W}$. While $\mathbf{W}$ may be compact, if we demand only the convergence of $X_j(w)$, which is the position at time $j < 0$, the functions $q(z|w)$ are not well defined on $\mathbf{W}$. So we need to construct a compactification $\overline{\mathbf{W}}$ of $\mathcal{W}$, such that, there is a natural map of $\overline{\mathbf{W}}$ onto $\overline{\mathcal{W}}$, which may not be one to one.

**1**. $\mathcal{W}$ is dense in $\overline{\mathbf{W}}$ and is unramified.

**2**.The functions $\{x_j(w)\}$, defined on $\mathbf{W}_n$ for $n \geq -j$, have continuous extensions, as maps into $Z^d$, to the closure of $\mathcal{W}$ in $\overline{\mathbf{W}}$ .

**3**.The number of visits and types of jumps $\{k_x(z,w)\}$ are continuous as maps into the space of integers with $\infty$ added i.e. into $Z^+ = \{0, 1, 2, \cdots, \infty\}$.

**4**.The transition probabilities $\{q(z|(T^*)^k w)\}, \{q(z|T_{z_k} \cdots T_{z_2} T_{z_1} w)\}$ are continuous maps for every $k \geq 0$ and every choice of $z, z_1, z_2, \cdots, z_k \in Z^d$.

**5**.The set $\overline{\mathcal{W}}$ while ramified, is embedded canonically as a subset of $\overline{\mathbf{W}}$ characterized by the property: for each $x \in Z^d$,

$$k_x(w) = \sum_{j \leq -1} \mathbf{1}_{\{x_j(w) = x\}}$$

In general only the inequality

$$k_x(w) \geq \sum_{j \leq -1} \mathbf{1}_{\{x_j(w) = x\}}$$

is true. Of course, if for some $x$, equality holds then for that $x$ and any $z$,

$$k_x(z, w) = \sum_{j \leq -1} \mathbf{1}_{\{x_j(w) = x, x_{j+1} = x+z\}}$$

We continue to call it $\overline{\mathcal{W}}$.

**6**.The maps $T^*$ and $\{T_z\}$ are continuous maps of $\overline{\mathbf{W}}$ into itself. (With $T^*$ we need to leave out $\mathbf{W}_0 = \{\phi\}$)

The existence of such a compactification, unique if it is minimal, is a routine exercise in real analysis. The compactification is only for convenience and will not solve any real problems. All it does is sweep some of the natural difficulties of the long range dependence in our model under the rug. We have to prove later that we do not have to look too closely underneath the rug.

For any $w \in \overline{\mathbf{W}}$ we denote its range $S(w)$ as the set

$$S(w) = \{x : k_x(w) > 0\}$$

$S(w)$ always contains 0. While $S(w)$ contains the actual range $\{x_j(w) : j \leq 0\}$ it may in general be larger. However for $w \in \overline{\mathcal{W}}$ they are the same.

We denote by $\overline{\mathcal{I}}$ the convex set of $T^*$ invariant probability measures on $\overline{\mathbf{W}}$ and by $\overline{\mathcal{E}}$ its extremals or the ergodic ones. For each $\mu \in \overline{\mathcal{I}}$, there is a natural process with stationary increments $z_j(w) = x_j(w) - x_{j-1}(w)$ and we can define its mean drift by

$$m(\mu) = E^{\mu}[z_0] = E^{\mu}[-x_{-1}]$$

**Lemma 1**: Suppose $\alpha \in \overline{\mathcal{E}}$ and $m(\alpha) \neq 0$. Let $s \in S^{d-1}$ be such that $< s, m(\alpha) >> 0$. Let $S(w)$ denote the support of $w$ defined earlier. If

$$\alpha\big[w : \sup_{x \in S(w)} < s, x >< \infty] > 0$$

then $\alpha$ is supported on $\overline{\mathcal{W}}$.

**Proof**: The proof is elementary. By ergodicity, the above probability, if it is not 0, is actually 1. Because $< s, m(\alpha) >> 0$ the whole scene is drifting away in the opposite direction and all the new visits are eventually to sites that are not in the current $S(w)$. That means eventually there are no excess visits other than the ones generated from the walk. Since the scene is invariant this was always true. In other words $\alpha[\overline{\mathcal{W}}] = 1$

**A combinatorial result.**

We start with $n$ steps $z_1, z_2, \cdots, z_n$ and let $X_n = z_1 + z_2 + \cdots + z_n$. This generates a history of $n$ walks, $\{w_j; 1 \leq j \leq n\}$ where $w_j = T_{z_j} \cdots T_{z_2} T_{z_1} \phi$ and $\phi$ is the starting point with no history. An empirical measures on $\mathcal{W}$ can be defined by

$$R_n = \frac{1}{n}[\sum_{j=0}^{n} \delta_{T_{z_j} \cdots T_{z_1} \phi}]$$

It is viewed as a measure on the compactification $\overline{\mathbf{W}}$. We assume that $\frac{X_n}{n} \to \xi \in R^d$. Let us denote by $\mathcal{C}$ the set of limit points of $R_n$ in the space of probability measures $\mathcal{M}(\overline{\mathbf{W}})$ on the compactified space $\overline{\mathbf{W}}$. We recall that $\overline{\mathcal{I}}$ is the set of $T^*$ invariant measures on $\overline{\mathbf{W}}$ and $\overline{\mathcal{E}}$ is the set of extremals in $\overline{\mathcal{I}}$. $\mathcal{E}$ was defined as the set of ergodic measures supported on $\overline{\mathcal{W}}$ and we denote by $\mathcal{E}_0$ measures $\alpha \in \mathcal{E}$ with $m(\alpha) \neq 0$. Clearly $\mathcal{E}_0 \subset \mathcal{E} \subset \overline{\mathcal{E}}$.

**Theorem 2.** Every $\mu \in \mathcal{C}$ is invariant under $T^*$. And as such, it has an integral representation over the ergodic measures (extremals)

$$\mu = \int_{\overline{\mathcal{E}}} \alpha \hat{\mu}(d\alpha)$$

$$\xi = \int_{\overline{\mathcal{E}}} m(\alpha) \hat{\mu}(d\alpha)$$

We can write $\hat{\mu} = \hat{\mu}_1 + \hat{\mu}_2$ in such a way that $\hat{\mu}_2$ is supported on $\mathcal{E}_0$ and

$$\xi = \int_{\mathcal{E}_0} m(\alpha) \hat{\mu}_2(d\alpha)$$

$$0 = \int_{\overline{\mathcal{E}}} m(\alpha) \hat{\mu}_1(d\alpha)$$

**Proof**: The theorem is essentially combinatorial in character. All the probabilities come from the empirical measures. The first part of the theorem is routine ergodic theory. Let us concentrate on the second part. If $\xi = 0$, there is nothing to prove and we can take $\hat\mu_2 = 0$. Let us therefore assume that $\xi \neq 0$. Let us denote by

$$B_\mu = \left\{ \int_{\mathcal{E}_0} \varphi(\alpha) m(\alpha) \hat\mu(d\alpha) : 0 \le \varphi \le 1 \right\}$$

$B_\mu \subset R^d$ is obviously a compact convex set in $R^d$ containing the origin. The problem is to show that $\xi \in B_\mu$. Assume that $\xi \notin B_\mu$. Then there is $s \in S^{d-1}$ and $a > a' > 0$ such that

(3) $$< s, \xi >> a >\ge \sup_{\eta \in B_\mu} < s, \eta >$$

We now consider the sequence of real numbers $y_j =< s, z_1 + z_2 + \cdots + z_j >$ and define successive times $\tau_1 \le \tau_2 \cdots \le \tau_j \le \tau_{j+1} \le \cdots$ as

$$\tau_j = n \wedge \inf\{i : y_i \ge j\}$$

These are times at which $y_i$ reaches successive predetermined levels that increase in unit steps. Notice that $\{y_j\}$ can jump over more than one level in a single step in which case $\tau_j = \tau_{j+1}$. There will be a last $\tau_r < n$ with $y_{\tau_r} \ge r$ and we will then have $y_n \le (r+1)$. We pick a (large) integer $k$, fix it, and call the interval $[\tau_j + 1, \tau_{j+1}]$ a 'good run' if $y_i \ge y_{\tau_j} - k$ for all $i \in [\tau_j + 1, \tau_{j+1}]$, and a 'bad run' otherwise. If $\tau_j = \tau_{j+1}$ the run is empty and it is natural to call it good. Each bad run uses up at least $\frac{k}{C}$ steps and therefore there can be atmost $\frac{nC}{k}$ bad runs. Each bad run gains a distance that is atmost $C+1$ and therefore the sum total of gains during bad runs is at most $\frac{nC(C+1)}{k}$. The last incomplete run may be negative, but in any case, is bounded above by 1. Therefore

(4) $$\sum_{\text{good runs}} (y_{\tau_{j+1}} - y_{\tau_j}) \ge n[a - \frac{C(C+1)}{k}] - 1$$

Note that during a good run, the walker currently at site $x$ has never before visited any site $x'$ with $< s, x' - x >> k+1$. The number of steps $j$, that are part of a good run, is at least $n(\frac{a}{C} - \frac{(C+1)}{k}) - \frac{1}{C}$. Let us denote by

$$\nu_{n,k} = \frac{1}{n} \sum_{j \in \text{'goodruns'}} \delta_{w_j}$$

the empirical distribution of $w_j$ over all the good runs, still normalized by $n$. The total mass of $\nu_{n,k}$ can be less than 1. It is at least $\frac{a}{C} - \frac{(C+1)}{k} + o(1)$. We now let $n \to \infty$, along a subsequence if necessary, and get a limit $\nu_k$. We assume that we have taken one subsequence that works for all $k$. The limiting measure $\nu_k$ may not be invariant because there are boundaries. As we let $k \to \infty$, $\nu_k$ is monotone and the limit $\nu$, which exists, is invariant. This is because although a proportion of $j$ among $[1, n]$ may be omitted as part of a bad run, the number of such runs is at most $n\frac{C(C+1)}{k}$. The contribution from the boundary is at most $\frac{2C(C+1)}{k}$ and is negligible for large $k$. We get in the limit a measure $\nu \leq \mu$ which may still have total mass less than 1, but will now be invariant. There is then a measure $\hat{\nu}$ on $\overline{\mathcal{E}}$ that represents it. Note that $\hat{\nu} \leq \hat{\mu}$. Since $\nu_{n,k}$ was constructed during 'good runs' we have for every $n$,

$$\nu_{n,k}\left[w : S(w) \cap \{x :< s, x >\geq k+1\} \neq \emptyset\right] = 0$$

which leads to

$$\nu_{k}\left[w : S(w) \cap \{x :< s, x >\geq k+1\} \neq \emptyset\right] = 0$$

Since the convergence of $\nu_k$ to $\nu$ is monotone

$$\|\nu_k - \nu\| = |\nu(\overline{\mathbf{W}}) - \nu_k(\overline{\mathbf{W}})| = \delta(k) \to 0$$

as $k \to \infty$. Therefore

$$\lim_{k\to\infty} \nu\left[w : S(w) \cap \{x :< s, x >\geq k+1\} \neq \emptyset\right] = 0$$

or

$$\sup_{x\in S(w)} < s, x >< \infty$$

a.e $\nu$. This forces

$$\hat{\nu}\left[\alpha : \alpha[\sup_{x\in S(w)} < s, x >< \infty] < 1\right] = 0$$

and in view of Lemma 1, we can conclude that $\hat{\nu}$ is supported on $\mathcal{E}$. Relation (4) on the other hand translates into

$$\int < s, m(\alpha) > \hat{\nu}(d\alpha) \geq a$$

after the limit on $n$ and $k$ are taken. Finally,

$$\int_{\mathcal{E}_0} < s, m(\alpha) > \hat{\nu}(d\alpha) = \int_{\mathcal{E}} < s, m(\alpha) > \hat{\nu}(d\alpha) \geq a$$

Because $\nu \leq \mu$, it follows that $\hat{\nu} \leq \hat{\mu}$ and $\hat{\nu}(d\alpha) = \varphi(\alpha)\hat{\mu}(d\alpha)$ for some $0 \leq \varphi \leq 1$. Hence

$$\sup_{0\leq\phi\leq 1} \int_{\mathcal{E}_0} < s, m(\alpha) > \phi(\alpha)\hat{\mu}(d\alpha) \geq a$$

which contradicts (3)

We will provide a proof of the large deviation upper bound. We will begin with some general facts regarding the local decay rate for Markov chaims on a countable sate space.

Let $\mathbf{X}$ be a countable state space and $\pi(x, y)$ the transition probability of a Markov chain on $\mathbf{X}$. We are interested in the local decay rate of the $n$-step transition probabilities $\pi^{(n)}(x, y)$. We will assume that the chain is irreducible with some period $p$. In addition we will assume that for each $x$, $\pi(x, \cdot)$ is supported on a finite set $F_x \subset \mathbf{X}$. For each $0 \leq \sigma < \infty$ we consider the convex set $\mathcal{U}_\sigma$ of positive sloutions of

$$\sum_y \pi(x, y)u(y) = e^{-\sigma}u(x)$$

normalized so that $u(x_0) = 1$ for some fixed point $x_0 \in \mathbf{X}$. In general $\mathcal{U}_\sigma$ may be empty. Note that the function $1 \in \mathcal{U}_0$. There exists $\lambda \geq 0$ which may be $+\infty$, such that $\mathcal{U}_\sigma$ is nonempty if and only if $\sigma \leq \lambda$. In addition for any fixed $x, y \in \mathbf{X}$

$$\limsup_{n \to \infty} \frac{1}{n} \log \pi^{(n)}(x, y) = -\lambda$$

If we let $n \to \infty$ along the subsequence dictated by the periodicity we have

$$\lim_{n \to \infty} \frac{1}{n} \log \pi^{(n)}(x, y) = -\lambda$$

Let $F$ be a finite set we define the exit time $\tau_F = \inf\{n : X_n \notin F\}$ and compute for $x \in F$ the expected value

$$u_F(\sigma, x) = E[e^{\sigma \tau_F} | X_0 = x]$$

which may or may not be finite. Let us define

$$\lambda_F = \Big\{ \sup \sigma : u_F(\sigma, x) < \infty \text{ for all } x \in F \Big\}$$

Then $\lambda = \inf_F \lambda_F$. The proofs of these assertions are elementary and depend on the fact that $u_F(\sigma, x)$ is the solution of

$$\sum_y \pi(x, y)u(y) = e^{-\sigma}u(x) \text{ for } x \in F$$

$$u(y) = 1 \text{ for } x \in F^c$$

For any finite set $F$, the value $\lambda_F$ can be characterized as $-\log \mathit{spect}\, \pi_F$ where $\pi_F$ is the substochastic matrix $\{\{\pi(x, y)\} : x, y \in F\}$ and *spec* refers to its spectral radius.

We next consider a simple dynamic programming problem. At each site $x$, a choice of $\pi(x, \cdot)$ can be made from a set $\mathcal{P}_x$ of possible distributions supported on some fixed finite set $N_x$. The goal is to maximize $u_F(\sigma, x)$. The Bellman equation for this problem is

$$\sup_{\pi(x, \cdot) \in \mathcal{P}_x} \sum_y \pi(x, y)u(y) = e^{-\sigma}u(x) \text{ for } x \in F$$

$$u(y) = 1 \text{ for } x \in F^c$$

which may or may not have a solution for a give $\sigma$. There will be a largest value $\widehat{\lambda}_F$ such that $u = \widehat{u}_F$ exists if $\sigma < \widehat{\lambda}_F$. It is clear from the Bellman equation that $\widehat{u}_F$ and $\widehat{\lambda}_F$ are not changed if we repalce each $\mathcal{P}_x$ by its convex hull. Moreover

$$\widehat{\lambda}_F = \inf_{\pi:\pi_x \in \mathcal{P}_x, x\in F} \lambda_F(\pi)$$

While the quenched rate function $h(\xi)$ is not in general explicit, $h(0)$ has a natural and important interpretation. Consider for each environment $\omega$ the $n$ step transition probability $\pi^{(n)}(\omega, 0, 0)$. From the Chapman-Kolmogorov equations, we have

$$\pi^{(n+m)}(\omega, x, x) = \sum_{y\in z^d} \pi^{(n)}(\omega, x, y)\pi^{(m)}(\omega, y, x) \geq \pi^{(n)}(\omega, x, x)\pi^{(m)}(\omega, x, x)$$

making $-\log \pi^{(n)}(\omega, x, x)$ subadditive. Therefore, if $p$ is the period

$$k(x, \omega) = -\lim_{n\to\infty} \frac{1}{np} \log \pi^{(np)}(\omega, x, x) = \inf_n [-\frac{1}{n} \log \pi^{(n)}(\omega, x, x)]$$

exists for every $x$ for almost all $\omega$. Moreover, by ellipticity $k(x, \omega) = k(\omega)$ is independent of $x$ and, by ergodicity, it is then equal to a constant $k = k(\beta)$. In particular

$$\pi^{(n)}(\omega, 0, 0) \leq e^{-k(\beta)\,n}$$

for every $n$, and $\omega$ in the support of $P$. The value $k(\beta)$ is the largest constant for which this is true. According to our earlier discussion, we can define $k(\beta)$ alternately in the following way. For each $x$, let us choose $\mathcal{P}_x$ as $\{p(\cdot - x)\}$ as $p(\cdot)$ varies over $\mathcal{K}$. Then, for a finite set $F$, the largest spectral radius over all possible choices of $\pi$ with $\pi(x, \cdot) \in \mathcal{P}_x$, is $e^{-\widehat{\lambda}_F}$ and

$$k(\beta) = \inf_F \widehat{\lambda}_F$$

depends only on the convex hull $\widehat{\mathcal{K}}$ of the support $\mathcal{K}$ of the distribution $\beta$ on $\mathcal{M}(Z^d)$.

Now we can integrate with respect to $P$ to obtain

$$\bar{Q}[x_n = 0] \leq e^{-k(\beta)\,n}$$

Of course $k(\beta)$ can and often is 0. The case $k(\beta) = 0$ is called the 'nestling' case and $k(\beta) > 0$ the 'non-nestling' case. There is a simple formula for $k(\beta)$.

**Lemma:** The value of $k(\beta)$ is given by

$$k(\beta) = -\log \big[ \inf_{\theta\in R^d} \sup_{p\in\widehat{\mathcal{K}}} [\sum_z e^{<\theta, z>} p(z)]\big]$$

where $\widehat{\mathcal{K}}$ is the convex hull of the support $\mathcal{K}$ of $\beta$. In order that $k(\beta) > 0$, it is necessary and sufficient that 0 be not in the closed convex hull of the support of the distribution of $m(p) = \sum_z zp(z)$ induced by the distribution $\beta$ on $\mathcal{M}(Z^d)$.

**Proof:** Suppose there is $\theta$ such that

$$\sup_{p\in\widehat{\mathcal{K}}}[\sum_z e^{<\theta,z>}p(z)] \le e^{-\lambda}$$

then, for any environment $\omega$, one has by induction

$$E^{P^\omega}\big[\exp[<\theta, x_n>]\big] \le e^{-n\lambda}$$

and this in turn implies immediately, that

$$\pi^{(n)}(\omega,0,0) \le e^{-n\lambda}$$

or $k(\beta) \ge \lambda$. On the other hand, because

$$\Psi(p(\cdot),\theta) = \log[\sum_z e^{<\theta,z>}p(z)]$$

is convex in $\theta$ and concave in $p(\cdot)$, by standard minimax argument

$$\inf_{\theta\in R^d}\sup_{p\in\widehat{\mathcal{K}}}[\sum_z e^{<\theta,z>}p(z)] = \sup_{p\in\widehat{\mathcal{K}}}\inf_{\theta\in R^d}[\sum_z e^{<\theta,z>}p(z)]$$

If there a $p(\cdot)\in\widehat{\mathcal{K}}$ such that

$$c = \inf_\theta \sum_z e^{<\theta,z>}p(z) \ge e^{-\lambda}$$

then, the uniform environment of simple random walk with jump distributed according to $p(\cdot)$ will have the property

$$\lim_{n\to\infty}\frac{1}{n}\log\pi^{(n)}(0,0) = \log c \ge -\lambda$$

establishing $k(\beta) \le \lambda$. This proves the first part of the lemma. Now, $k(\beta) = 0$ if and only if there is a $p(\cdot)\in\widehat{\mathcal{K}}$ such that

$$\inf_{\theta\in R^d}[\sum_z e^{<\theta,z>}p(z)] = 1$$

i.e. $\sum_z z\,p(z) = 0$.

Since we have a Feller process with transition probabilities $q(w,w')$ on the compact space $\overline{\mathbf{W}}$ the standard theory of large deviation applies and we get a lower semi-continuous rate function $\mathcal{J}(\mu)$ defined by

$$\mathcal{J}(\mu) = \sup_{u\in C^+(\overline{\mathbf{W}})}\int \log\frac{u(w)}{(qu)(w)}\mu(dw)$$

which is infinite unless $\mu \in \widehat{\mathcal{I}}$, the set of invariant measures on $\overline{\mathbf{W}}$. For the empirical measure $\mathcal{R}_n$ we get the estimates

$$\limsup_{n\to\infty} \frac{1}{n} \sup_w \log Q^w \big[\mathcal{R}_n \in C\big] \le - \inf_{\mu\in C} \mathcal{J}(\mu)$$

for closed sets $C$, and for open sets $G$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \sup_w \log Q^w \big[\mathcal{R}_n \in G_\mu\big] \ge - \inf_{\mu\in G\cap\widehat{\mathcal{E}}} \mathcal{J}(\mu)$$

which is much weaker and can only be improved under some sort of transitivity condition.

**Lemma:** With $k = k(\beta)$ defined before

$$\inf_{\substack{\mu\in\widehat{\mathcal{I}}\\ m(\mu)=0}} \mathcal{J}(\mu) \ge k(\beta)$$

**Proof:** We need to find a $u$ and use the formula

$$\mathcal{J}(\mu) \ge \int \log \frac{u(w)}{(qu)(w)} \mu(dw)$$

We will try $u(w) = e^{<\theta, z_0(w)>}$. Since $m(\mu) = 0$, we obtain the lower bound

$$\mathcal{J}(\mu) \ge - \inf_{\theta\in R^d} \int \log[\sum_z q(z|w) e^{<\theta,z>}]\mu(dw)$$

Since $q(w,\cdot) \in \mathcal{K}$, we are done.

**Lemma** For $0 \ne \xi \in R^d$ and any open set $G \ni \xi$

$$\liminf_{n\to\infty} \bar{Q}[\frac{x_n}{n} \in G] \ge -H(\xi)$$

**Proof:** Let $\alpha \in \mathcal{E}$ with $m(\alpha) = \xi$ be given. Let us denote by

$$\hat{\alpha}(z|w) = \alpha[z_1 = z|w]$$

the conditional probability of the next step being $z$ given the past. Then

$$I(\alpha) = E^\alpha[\sum_z \hat{\alpha}(z|w) \log \frac{\hat{\alpha}(z|w)}{q(z|w)}]$$

By the ergodic theorem, for any $\epsilon > 0$, as $n \to \infty$

$$\alpha\Big[|\bar{z}(n) - \xi| \le \epsilon, |\bar{I}(n) - I(\alpha)| \le \epsilon\Big] \to 1$$

where

$$\bar{z}(n) = \frac{1}{n}\sum_{i=1}^{n} z_i$$

$$\bar{I}(n) = \frac{1}{n}\sum_{i=1}^{n} \log\frac{\hat{\alpha}(z_i|w_{i-1})}{q(z_i|w_{i-1})}$$

Moreover we can assume that $\ell$ is large enough that for all $n$,

$$\alpha[H(n, S(w), \mathbf{z}) \le \ell] \ge \frac{1}{2}$$

for some $s \in S^{d-1}$ with $< s, \xi >> 0$. We wish to estimate

$$\begin{aligned}
&Q^{\phi}[|\bar{z}(n) - \xi| \le \epsilon] \\
&\quad \ge Q^{\phi}[|\bar{z}(n) - \xi| \le \epsilon, H(n, S(w), \mathbf{z}) \le \ell] \\
&\quad \ge e^{-C\ell} \sup_{w} P^{w}[|\bar{z}(n) - \xi| \le \epsilon, H(n, S(w), \mathbf{z}) \le \ell] \\
&\quad \ge e^{-C\ell} \int P^{w}[|\bar{z}(n) - \xi| \le \epsilon, H(n, S(w), \mathbf{z}) \le \ell]\alpha(dw) \\
&\quad = e^{-C\ell} \int_{\{|\bar{z}(n)-\xi|\le\epsilon\}\cap\{H(n,S(w),\mathbf{z})\le\ell\}} R_n(w)\alpha(dw) \\
&\quad \ge e^{-C\ell} e^{-n(I(\alpha)+\epsilon)} \alpha\big[|\bar{z}(n) - \xi| \le \epsilon, |\bar{I}(n) - I(\alpha)| \le \epsilon, H(n, S(w), \mathbf{z}) \le \ell\big]
\end{aligned}$$

**Theorem:** The function

$$H(\xi) = \inf_{\substack{\mu\in\mathcal{E} \\ m(\mu)=\xi}} \mathcal{J}(\mu)$$

is a convex function in every half space $\{\xi :< s, \xi >> 0\} \subset R^d$. The limit $\lim_{\xi\to 0} H(\xi)$ exists and equals $k(\beta)$. If we define $H(0)$ as this limit then $H(\cdot)$ is a convex function on $R^d$.

**Proof:** First, assume that $\xi = a\xi_1 + (1 - a)\xi_2$, with both $\xi_1$ and $\xi_2$ belonging to the same half space, i.e. there exists $s \in S^{d-1}$ such that $< s, \xi_i >> 0$ for $i = 1, 2$. There are processes with stationary increments $\mu_i$ that are ergodic with means $\xi_i$ and $\mathcal{J}(\mu_i) = c_i$. The measure $\mu = a\mu_1 + (1 - a)\mu_2$ is not ergodic but has stationary increments. It can be approximated by ergodic ones by the usual method of taking a block of length $\ell$ and constructing a product measure over distinct blocks and then averaging over $\ell$ translates to produce some thing that is ergodic and stationary. If we call this approximation by $\mu_\ell$, then it is ergodic and $m(\mu_\ell) = m(\mu)$ for every $\ell$. One needs to prove only that

$$\lim_{\ell\to\infty} \mathcal{J}(\mu_\ell) = \mathcal{J}(\mu)$$

From the lower semi-continuity of $\mathcal{J}(\cdot)$, it is clearly enough to prove that $\mu_\ell \to \mu$ weakly as measures on $\overline{\mathbf{W}}$. From our earlier results it follows that the limit $\mu'$ of $\mu_\ell$ along any subsequence will be represented by

$$\mu' = \theta \int_{\mathcal{E}} \alpha\nu'(d\alpha) + (1 - \theta) \int_{\widehat{\mathcal{E}}} \alpha\nu(d\alpha)$$

with

$$\int_{\mathcal{E}} m(\alpha)\nu'(d\alpha) = 0$$

However, $\mu'$ on $\overline{\mathbf{W}}$ projects to $\mu$ on $\mathbf{W}_{\infty}^{tr}$. Therefore both $\nu$ and $\nu'$ can only be supported on measures $\alpha$ with $m(\alpha) = \xi_1$ or $\xi_2$. Since 0 is not in the convex hull of $\xi_1$ and $\xi_2$ we must have $\nu' = 0$. Therefore $\mu' = \mu$ and we are done.

Now we turn to the proof of $\lim_{\xi\to 0} H(\xi) = k(\beta)$. If $m(\mu_n) = \xi_n$ and $\xi_n \to \xi$, then along a subsequence $\mu_n \to \mu$ and $m(\mu) = 0$ even if it is not ergodic. By the semicontinuity of $\mathcal{J}(\cdot)$

$$\mathcal{J}(\mu) \le \liminf_{n\to\infty} \mathcal{J}(\mu_n) \le \liminf_{n\to\infty} H(\xi_n)$$

But from Lemma

$$k(\beta) \le \inf_{m(\mu)=0} \mathcal{J}(\mu)$$

Hence

$$k(\beta) \le \liminf_{\xi\to 0} H(\xi)$$

To prove the inequality in the reverse direction, given $\xi \in S^{d-1}$, for sufficiently small $\delta$, we have to construct an ergodic $\alpha \in \mathcal{E}$ with $m(\alpha) = \delta\xi$ and $\mathcal{J}(\alpha) \le k(\beta) + \epsilon(\delta)$, where $\epsilon(\delta) \to 0$ with $\delta$. This can be accomplished as follows. Let $\xi \in S^{d-1}$, integers $L, k$, and $\delta > 0$ be given. Let us select a deterministic path connecting the origin and a lattice site close to $k\delta\xi$ in $Ck\delta$ steps, where each step has conditional probability atleast $\rho$. Assume that the path is such that the end point is a boundary point of a cube $C_L$ of size $L$ and the path has no other common points with the cube $C_L$. We will construct a probability distribution $\mu_{L,k}$ of a walk that will start at where the deterministic path ends, remain inside the cube $C_L$ for $k(1 - C\delta)$ steps and ends up at the boundary point where it entered, making essentially a loop. The two walks together will produce a probability distribution, which we will denote by $\mu_{L,\delta,k,\xi}$ which in $k$ steps goes from 0 to $k\delta\xi$. We can take a countable product of this and will get a process whose increments will be stationary under the $k$ shift. We can do the averaging over $k$ steps and get an ergodic process with stationary increments, that we call $\alpha_{L,\delta,k,\xi}$. It is not difficult to estimate $\mathcal{J}(\alpha_{L,\delta,k,\xi})$. The deterministic part contributes at most $Ck\delta \log \frac{1}{\rho}$ and the part confining the walk to a box of size $L$ depends on the choice of $\mu_{L,k}$ and it can be chosen so that it contributes at most $\sigma_L k(1 - C\delta) + o(k)$, where $\sigma_L$ is given by

$$\sigma_L = -\liminf_{n\to\infty} \frac{1}{n} \log Q^{\phi}\Big[z_1 + z_2 + \cdots + z_j \in C_L \text{ for all } 1 \le j \le n\Big]$$

and $\sigma_L \to k(\beta)$ as $L \to \infty$. this leads to an estimate of the form

$$\mathcal{J}(\alpha_{L,\delta,k,\xi}) \le \sigma_L(1 - C\delta) + C\delta \log \frac{1}{\rho} + o(1)$$

as $k \to \infty$. We let $k \to \infty$ and then $L$ to infty, to obtain the estimate

$$H(\delta\xi) \le k(\beta)(1 - C\delta) + C\delta \log \frac{1}{\rho}$$

for arbitrary $\xi \in S^{d-1}$ and $\delta$ small enough.

Finally to prove convexity at 0, we need to show that for any $\{\xi_j\}$ with convex combination $\sum_j a_j \xi_j = 0$ we have $\sum_j a_j k(\xi_j) \geq k(\beta)$. But $\mu = \sum_j a_j \mu_j$ with $m(\mu_j) = \xi_j$ has $m(\mu) = 0$ and therefore

$$\sum_j a_j \mathcal{J}(\mu_j) = \mathcal{J}(\mu) \geq k(\beta)$$

and we are done.

**Proof of Theorem 1.** We have already established the lower bounds and the upper bound at 0. We need to prove only the upper bound for $\xi \neq 0$. In view of Theorem 2, we have the following inequalities valid for closed sets $C$ not containing 0.

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log Q^{\phi}\big[\frac{x_n}{n} \in C\big] &\leq - \inf_{\xi \in C} \inf_{\theta} \left[(1-\theta)k(\beta) + \theta H(\frac{\xi}{\theta})\right] \\
&\leq - \inf_{\xi \in C} \inf_{\theta} \left[(1-\theta)H(0) + \theta H(\frac{\xi}{\theta})\right] \\
&\leq - \inf_{\xi \in C} H(\xi)
\end{aligned}$$