PRINCÍPIO DE GRANDES DESVIOS EM CADEIAS DE MARKOV FINITAS E HOMOGÊNEAS

Marco Antônio GIACOMELLI[1]

▪ RESUMO: O princípio de Grandes Desvios (PGD) de Cramer-Chernoff para seqüências de variáveis aleatórias independentes e identicamente distribuídas surgiu em 1952. Este princípio teve generalizações para variáveis aleatórias com algum tipo de dependência estocástica. O artigo aqui proposto se restringe ao contexto de Cadeias de Markov finitas e homogêneas. Inicialmente são dados alguns resultados e definições que serão utilizados posteriormente. O objetivo é analisar três situações onde são aplicados PGD: a média amostral, a medida empírica e a medida empírica aos pares. A referência central é Ellis (1985), cuja nomenclatura são somas de nível 1, nível 2 e nível 3 para a média, medida empírica e medida empírica aos pares, respectivamente.

▪ PALAVRAS-CHAVE: Princípio de Grandes Desvios, média amostral em Cadeias de Markov, medida empírica em Cadeias de Markov.

1 Introdução

As primeiras idéias sobre Grandes Desvios estão direcionadas a seqüências de variáveis aleatórias (v.a's) independentes e identicamete distribuídas (i.i.d), assumindo valores em **R** e com esperança finita. Mas o que são Grandes Desvios? A resposta mais simples é que a Teoria de Grandes Desvios concentra-se em

[1]Departamento de Estatítica da Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, e-mail:giacomo@mat.ufrgs.br

determinados eventos raros, ou seja, certa classe de eventos cuja probabilidade tende a zero. Esta probabilidade tende a zero, por exemplo, em função do tamanho da amostra. O exemplo clássico refere-se a uma seqüência $\{X_n\}_{n\geq 1}$ de v.a's i.i.d. Sabe-se, pela Lei dos Grandes Números (LGN), que para $\forall\,\epsilon > 0$, $\lim_{n\to\infty} P(|\bar{X}_n - m| \geq \epsilon) = 0$, sendo $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média amostral e $m = EX_1 < \infty$. A interpretação deste limite é que para n muito grande a ocorrência do evento $[|X_n - m| \geq \epsilon]$ é muito pouco provável, isto é, a probabilidade deste evento tende a zero à medida que o tamanho da amostra aumenta. Assim, diz-se que $[|X_n - m| \geq \epsilon]$ é um evento do tipo "Grande Desvio"com relação à LGN.

No que segue veremos um breve histórico sobre a teoria de Grandes Desvios. Cramer (1937) e Chernoff (1952) estabeleceram os primeiros resultados sobre Grandes Desvios, especificamente no contexto de v.a's i.i.d com esperança m finita e função geradora de momentos $M(t) = Ee^{tX_1}$ também finita. O Princípio de Grandes Desvios (PGD) de Cramer-Chernoff pode ser enunciado da seguinte maneira:

$$\liminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in G) \geq -\inf_{x\in G} \lambda(x), \text{ para } G \text{ aberto,} \tag{1.1}$$

$$\limsup_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in F) \leq -\inf_{x\in F} \lambda(x), \text{ para } F \text{ fechado,} \tag{1.2}$$

sendo $\lambda(x) \equiv \sup_{t\in\mathbf{R}} \{tx - \ln M(t)\}$, $x \in \mathbf{R}$, uma função convexa e não negativa denominada de transformada de Cramer. Observe que quando $\inf_{x\in int(G)} \lambda(x) = \inf_{x\in\bar{F}} \lambda(x)$, sendo $int(G)$ o interior de G e $\bar{F}$ o fecho de F, então por (1.1) e (1.2) temos

$$\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in F) = -\inf_{x\in F} \lambda(x)\,. \tag{1.3}$$

Somente quando (1.3) é satisfeita então dizemos que existe uma "equivalência logarítmica"entre $P(\bar{X}_n \in F)$ e $e^{-n\inf_{x\in F}\lambda(x)}$. A função λ, que satisfaz (1.1) e (1.2), é denominada de "função taxa"e também é conhecida por transformada de Legendre. A interpretação da equivalência logarítmica é que para $m \notin F$, $\ln P(\bar{X}_n \in F)$ tende para $-\infty$, quando $n \to \infty$, na mesma intensidade do que $-n\lambda(x)$ tende para $-\infty$. Observe que em alguns textos aparece a expressão: "a probabilidade de um evento do tipo Grande Desvio decai exponencialmente a zero". Esta expressão está se referindo à equivalência logarítmica, ou seja, o limite (1.3).

Neste momento é importante comentar a relação entre LGN e PGD. Podemos perceber que um PGD complementa a informação da LGN, isto é, além de sabermos que a probabilidade do evento $[\bar{X}_n \in F]$, $m \notin F$, tende a zero, o limite (1.3) nos diz como é este decaimento. Já, se $m \in F$, pela LGN $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in F) = 0$ e pelo PGD, $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in F) = -\lambda(m) = 0$, sendo que esta última igualdade decorre da propriedade de que $\lambda(m) = 0$ para m finita.

Posteriormente aos resultados de Cramer-Chernoff generalizações foram obtidas basicamente em duas direções: quanto ao espaço onde as v.a's estão definidas e quanto à existência de algum tipo de dependência estocástica entre as v.a's. Referente ao espaço das v.a's, podemos citar Lanford (1971), que considerou o espaço $\mathbf{R}^d$. Por outro lado, Donsker & Varadhan (1975a), fizeram generalizações para espaços mais gerais: Banach, Hausdorff e espaços infinito-dimensionais.

Sanov (1961) generalizou os trabalhos de Cramer-Chernoff estabelecendo um PGD para a medida empírica de v.a's i.i.d. Já, Gärtner (1977) estabeleceu um PGD para Cadeias de Markov finitas e Donsker & Varadhan (1975b, 1976 e 1983) estabeleceram um PGD para a medida empírica em um contexto mais geral, isto é, em Processos de Markov.

A partir das décadas de 70 e 80 surgiram novos resultados sobre Grandes Desvios em áreas tais como: processos de difusão (Freidlin & Wentzell, 1984), e Mecânica Estatística (Varadhan, 1984), Ellis (1985) e outros autores. No Brasil uma importante contribuição sobre Grandes Desvios é devida a Vares (1985).

Neste artigo nos concentraremos em Cadeias de Markov irredutíveis, finitas, aperiódicas e estacionárias, sendo que iremos estabelecer um PGD para soma de nível 1, ou seja, a média amostral, para soma de nível 2, que é a medida empírica, e para soma de nível 3, conhecida como medida empírica aos pares. Esta nomenclatura de somas de níveis 1,2 e 3 é utilizada por Ellis (1985) e será devidamente definida adiante.

2 Alguns resultados e definições

Nesta seção serão definidos alguns conceitos que serão posteriormente utilizados. Também serão enunciados alguns resultados importantes para estabelecer um PGD em Cadeias de Markov.

Definição 2.1 *Seja* $\{X_n\}_{n\geq 0}$ *uma Cadeia de Markov com espaço de estados* $S = \{1, \cdots, r\}$ *e matriz de transição* $P = \{p(i,j)\}_{i,j\in S}$. *A Cadeia é dita irredutível se para qualquer par* $(i,j) \in S$ *existe* $m = m(i,j)$ *tal que* $p^m(i,j) > 0$, *sendo* $p^m(i,j)$ *a probabilidade de transição do estado* i *para* j *em* m *transições .*

Definição 2.2 *Uma Cadeia de Markov é dita aperiódica se o m.d.c* $\{n : p^n(i,j) > 0\} = 1$, $\forall\, i \in S$, *caso contrário ela é dita periódica* .

Definição 2.3 *Seja* μ_0 *a distribuição inicial de uma Cadeia de Markov e* $\mu_n = \mu_0 P^n$ *a distribuição de probabilidade na* $n-$*ésima transição. A cadeia é dita ergódica se* $\lim_{n\to\infty} \mu_n = \mu$, *independentemente de* μ_0.

O teorema a seguir é necessário para estabelecer propriedades algébricas das matrizes de transição e resultados sobre Grandes Desvios em Cadeias de Markov. De uma maneira geral, o teorema vale para matrizes não necessariamente estocásticas.

Teorema 2.1 *Seja a matriz* $B = \{b(i,j)\}_{i,j=1}^{r}$ *com entradas não negativas. Assuma que para qualquer par* (i,j) *existe* $m = m(i,j)$ *tal que* $b^m(i,j) > 0$, *sendo* $b^m(i,j)$ *a entrada* (i,j) *da* $m-$*ésima potência de* B. *Então* B *possui um autovalor* λ, *denominado de autovalor de Perron-Frobenius, tal que*:
(i) $\lambda > 0$ é real;
(ii) para qualquer outro autovalor θ de B, $|\theta| \leq \lambda$;
(iii) existem autovetores μ à direita e ν à esquerda, correspondentes a λ, com coordenadas estritamente positivas;
(iv) μ e ν são únicos a menos de uma constante;
(v) para todo $i = 1, \cdots, r$ e todo $\phi = (\phi(1), \cdots, \phi(r))$, com $\phi(j) > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \ln \left(\sum_{j=1}^{r} b^n(i,j)\phi(j) \right) = \lim_{n\to\infty} \frac{1}{n} \ln \left(\sum_{j=1}^{r} \phi(j) b^n(j,i) \right) = \ln \lambda.$$

Prova (i), (ii), (iii) e (iv) podem ser vistos em Seneta (1981). O item (v) pode ser encontrado em Dembo & Zeitouni (1993). □

Teorema 2.2 (Teorema Ergódico para Cadeias de Markov) *Seja uma Cadeia de Markov irredutível e aperiódica. Então*:
(i) $\lim_{n\to\infty} p^n(i,j) = \lim_{n\to\infty} p^n(j,j) = \mu(j)$ para $\forall\, i, j \in S$;
(ii) $\mu = (\mu(1), \cdots, \mu(r))$ é tal que $\mu(j) > 0$, $\mu(j) = \sum_{i\in S} \mu(i)p(i,j)$ e $\sum_{j\in S} \mu(j) = 1$;
(iii) a distribuição de probabilidade μ, em (i) e (ii), é única e é denominada de "medida invariante" da Cadeia de Markov.

Prova Há algumas maneiras de se provar este teorema. Citaremos duas. A primeira é através da Equação Discreta da Renovação, ver Karlin & Taylor (1975). A segunda é através da "Entropia Relativa", ver Liggett (1985). Nesta segunda maneira, utiliza-se o Teorema 2.1. □

Comentário 2.1 Na verdade, o Teorema 2.2 também vale para cadeias com uma única subclasse fechada e aperiódica.

Comentário 2.2 Quando a Cadeia de Markov for irredutível ou possuir uma única subclasse fechada, mas tiver período d, então temos a seguinte versão do Teorema 2.2:
(i) $\lim_{n\to\infty} p^{nd}(i,j) = \lim_{n\to\infty} p^{nd}(j,j) = d\mu(j)$, para $\forall\, i, j \in S$;
(ii) $\mu = (\mu(1), \cdots, \mu(r))$ é tal que $\mu(j) > 0$, $\mu(j) = \sum_{i\in S} \mu(i)p(i,j)$ e $\sum_{j\in S} \mu(j) = 1$;
(iii) μ, em (i) e (ii), é única.

Definição 2.4 (Somas de nível 1) *A soma de ordem n de nível 1 para uma Cadeia de Markov $\{X_n\}_{n\geq 0}$ é definida como*:

$$\bar{X}_n(\omega) \equiv \sum_{i=1}^{n} \frac{X_i(\omega)}{n}$$

e também denominada de "média amostral".

Comentário 2.3 Por simplicidade, também denotaremos a média amostral como $\bar{X}_n$, isto é, ficará implícito que existe um espaço amostral no qual a Cadeia está definida.

Definição 2.5 (Somas de nível 2) *A soma de ordem n de nível 2, também denominada de medida empírica para uma Cadeia de*

Markov, é definida como $L_n(\omega) \equiv (L_n(\omega, 1), \cdots, L_n(\omega, r))$*, sendo*

$$L_n(\omega, i) = \sum_{k=1}^{n} \delta_{X_k(\omega)}(i) \quad \text{e} \quad \delta_{X_k(\omega)}(i) = \begin{cases} 1, & \text{se} \quad X_k(\omega) = i \\ 0, & \text{caso contrário} \end{cases}, \; i = 1, \cdots, r.$$

Comentário 2.4 Se ω variar então $L_n(\cdot, j)$ é uma v.a. e para ω fixo $L_n(\omega, \cdot)$ é uma medida de probabilidade em $\mathcal{M}_\infty(\mathcal{S}) \equiv \left\{\nu : \sum_{\rangle=\infty}^{\nabla} \nu(\rangle) = \infty,\ \nu(\rangle) \geq \prime\right\}$.

Comentário 2.5 O vetor L_n tem como interpretação a freqüência relativa dos estados que a Cadeia assume em n observações.

Teorema 2.3 (Lei dos Grandes Números para Cadeias de Markov) *Considere uma Cadeia de Markov irredutível e aperiódica e seja* $\mu = (\mu(1), \cdots, \mu(r))$ *sua medida invariante. Então, para qualquer distribuição inicial* $\mu_0 = (\mu_0(1), \cdots, \mu_0(r))$*, tem-se :*
(i) $\lim_{n\to\infty} E_j\left(L_n(i)\right) = \mu(i)$, $i \in S$, onde E_j denota a esperança condicionada ao estado inicial $j \in S$;
(ii) $\forall\, \epsilon > 0$, $\lim_{n\to\infty} P_j\left(|L_n(i) - \mu(i)| > \epsilon\right) = 0$, $\forall\, i, j \in S$, onde P_j denota a probabilidade do evento $[|L_n(i) - \mu(i)| > \epsilon]$ condicionada ao estado inicial $j \in S$.
Prova ver Kemmeny & Snell (1960). □

Comentário 2.6 Para uma Cadeia de Markov com uma única subclasse fechada F, mas com período d, a versão do Teorema 2.3 fica sendo:
(i) $\lim_{n\to\infty} E_j\left(L_{nd}(i)\right) = d\mu(i)$, $i, j \in F$;
(ii) $\forall\, \epsilon > 0$, $\lim_{n\to\infty} P_j\left(|L_{nd}(i) - d\mu(i)| > \epsilon\right) = 0$, $\forall\, i, j \in F$.

As somas de nível 2 são generalizadas considerando freqüências relativas de pares de variáveis da Cadeia de Markov $\{X_n\}_{n\geq 0}$ com espaço de estados $S = \{1, \cdots, r\}$. Defina

$$Y_k = \begin{cases} (X_k, X_{k+1}), & \text{se} \;\; k \in 0, 1, \cdots, n-1 \\ (X_n, X_0), & \text{se} \;\; k = n \end{cases}.$$

Sendo assim, a seqüência $\{Y_k\}_{k\geq 0}$ assume valores em $S^2 = S \times S$.

Definição 2.6 (Somas de nível 3) *A soma de ordem* n *de nível 3, também denominada de medida empírica aos* *pares, para uma Cadeia de Markov, é definida como* $M_n \equiv (M_n(1,1), M_n(1,2), \cdots, M_n(r,r))$*, sendo*

$$M_n(i,j) = \sum_{k=1}^{n} \frac{\delta_{Y_k}(i,j)}{n} \quad \text{e} \quad \delta_{Y_k}(i,j) = \begin{cases} 1, & \text{se } Y_k = (i,j) \\ 0, & \text{caso contrário} \end{cases}, \ (i,j) \in S^2.$$

Comentário 2.7 O vetor M_n tem como interpretação a freqüência relativa dos pares (i,j). Observe que a medida M_n é consistente com a medida empírica L_n, isto é, $L_n(i) = \sum_{j=1}^{r} M_n(i,j) = \sum_{k=1}^{r} M_n(k,i)$.

Definição 2.7 (Princípio de Grandes Desvios) *Seja* $\{W_n\}_{n\geq 1}$ *uma seqüência de v.a's definidas em* $\{\Omega_n, \mathcal{F}_n, \mathcal{P}_n\}_{n\geq 1}$*, onde* Ω_n *é o espaço amostral,* $\mathcal{F}_n$ *a* $\sigma-$*álgebra e* P_n *uma medida de probabilidade. Dizemos que* $\{W_n\}_{n\geq 1}$ *satisfaz um princípio de Grandes Desvios, com função taxa* I*, se para a seqüência* $\{a_n : a_n > 0,\, n \geq 1\}$*, com* $\lim_{n\to\infty} a_n = +\infty$ *:*
(i) $\liminf_{n\to\infty} \frac{1}{a_n} \ln P_n\left(\frac{W_n}{a_n} \in G\right) \geq -\inf_{z\in G} I(z)$, G aberto;
(ii) $\limsup_{n\to\infty} \frac{1}{a_n} \ln P_n\left(\frac{W_n}{a_n} \in F\right) \leq -\inf_{z\in F} I(z)$, F fechado.

Definição 2.8 (Função convexa) *Uma função* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *é dita convexa se* f *não assume* $-\infty$*, é finita pelo menos para algum* $x \in \mathbf{R}$ *e* $\forall\, x, y \in \mathbf{R}$*,* $\forall\, \alpha \in [0,1]$*,* $f\left((1-\alpha)x + \alpha y\right) \leq (1-\alpha) f(x) + \alpha f(y)$ *.*

Definição 2.9 (Função semicontínua inferiormente) *Uma função* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *é dita semicontínua inferiormente, em* $a \in \mathbf{R}$*, se* $\forall \epsilon > 0$*,* $\exists\, \delta > 0$*, tal que para* $x \in \mathbf{R}$*, com* $|x - a| < \delta$*, tem-se* $f(a) - \epsilon < f(x)$ *.*

Definição 2.10 (Função taxa) *Uma função* $I : \mathbf{R} \longrightarrow [0, +\infty]$ *é dita "função taxa"se*:
(i) I não é identicamente $+\infty$;
(ii) I é semicontínua inferiormente ;
(iii) para $\forall L \in [0, +\infty)$ o conjunto $\{x : I(x) \leq L\}$ é compacto.

Definição 2.11 *Seja* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *convexa. Então* f *é dita fechada se* $\big\{(x,y) \in \mathbf{R^2} : y \geq f(x)\big\}$ *é conjunto fechado em* $\mathbf{R^2}$ *.*

Comentário 2.8 Uma função convexa é fechada se esta for semicontínua inferiormente.

Definição 2.12 *Seja* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$*. Definimos o domínio efetivo de* f *por* $dom(f) = \{x \in \mathbf{R} : f(x) < \infty\}$.

Proposição 2.1 *Se* f *é convexa e* $x \in int(dom(f))$ *então existem as derivadas laterais* $f'_{-}(x)$ *e* $f'_{+}(x)$ *.*
Prova ver Lima (1982). $\square$

Comentário 2.9 Da proposição acima segue que f restrita ao $int(dom(f))$ é contínua.

Definição 2.13 *Seja* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *uma função convexa. Para* $x \in \mathbf{R}$*, definimos a subdiferencial de* f *em* x *como*:

$$\partial f(x) = \{\alpha \in \mathbf{R} : f(y) \geq f(x) + \alpha(y - x), \forall y \in \mathbf{R}\}.$$

Um ponto de $\partial f(x)$ é dito subgradiente de f em x.

Proposição 2.2 *Para* f *função convexa,* $x \in int(dom(f))$ *se e somente se* $\partial f(x) = \left[f'_{-}(x), f'_{+}(x)\right]$.
Prova Ellis (1985) e Acker & Dickstain (1983). $\square$

Definição 2.14 *Seja* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *convexa. Definimos a transformada de Legendre* $f^* : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *como* $f^*(y) = \sup_{x \in \mathbf{R}} \{xy - f(x)\}$.

Teorema 2.4 *Seja* $f : \mathbf{R} \longrightarrow \mathbf{R} \cup \{+\infty\}$ *convexa e fechada. Então*:
(i) f^* é convexa e fechada;
(ii) $y \in \partial f(x)$ se e somente se $x \in \partial f^*(y)$;
(iii) $xy = f(x) + f^*(y)$ se e somente se $y \in \partial f(x)$;
(iv) $(f^*)^* = f$, ou seja, $f(x) = sup_{y \in \mathbf{R}} \{xy - f^*(y)\}$, para todo $x \in \mathbf{R}$.
Prova Ellis (1985). $\square$

Comentário 2.10 Pela Definição 2.8 f convexa não assume $-\infty$ e também é finita pelo menos para um $x \in \mathbf{R}$. Logo, $f^*(y) > -\infty$, ou seja, f^* não assume $-\infty$.

Proposição 2.3 *Seja* f *convexa e finita, portanto, fechada e derivável. Então*:
(i) f' é não decrescente;
(ii) $int(dom(f^*)) \subset imagem(f') \subset dom(f^*)$;

(iii) $\exists\, x_0$ tal que $y = f^{'}(x_0)$, isto é, $f^*(y) = x_0 y - f(x_0)$.
Prova
(i) Lima (1982);
(ii) Ellis (1985);
(iii) Defina $g_y(x) = xy - f(x)$. Então, $\frac{d^2}{dx^2} g_y = -f^{''}(x) \leq 0$, pois f é convexa. Lembre que se f é convexa então $f^{''} \geq 0$). Assim, a solução de $\frac{d}{dx} g_y = 0$ fornece um ponto de máximo, isto é, $y - f^{'}(x) = 0$ resulta que $y = f^{'}(x_0)$. Logo, $f^*(y) = yx_0 - f(x_0)$.

Definição 2.15 (Medida absolutamente contínua) *Sejam* $(\Omega, \mathcal{F})$ *um espaço mensurável,* ν *e* μ *medidas de probabilidade neste espaço. Dizemos que* ν *é absolutamente contínua, com relação a* μ*, se para cada* $A \in \mathcal{F}$*,* $\mu(A) = 0$ *implica em* $\nu(A) = 0$*. Denotamos por* $\nu \ll \mu$*.*

Definição 2.16 *Assuma* $0 \ln(0) \equiv 0$ *e* $0 \ln\left(\frac{0}{0}\right) \equiv 0$*. Então*:
(i) a entropia de um vetor de probabilidade ν é

$$H(\nu) \equiv -\sum_{i=1}^{r} \nu(i) \ln \nu(i);$$

(ii) a entropia relativa de um vetor de probabilidade ν, com respeito a outro vetor de probabilidade μ, é

$$H(\nu|\mu) \equiv \sum_{i=1}^{r} \nu(i) \ln\left(\frac{\nu(i)}{\mu(i)}\right).$$

Comentário 2.11 A entropia relativa tem como interpretação o grau de afastamento do vetor ν em relação a μ. De fato, $H(\nu|\mu) = 0$ somente quando $\nu = \mu$.

Teorema 2.5 *A entropia relativa, da Definição 2.16, é*:
(i) convexa e contínua em ν;
(ii) não negativa;
(iii) finita.
Prova
(i) Defina $\phi(x) = x \ln x$. Assim, $H(\nu|\mu) = \sum_{i=1}^{r} \mu(i) \phi\left(\frac{\nu(i)}{\mu(i)}\right)$. Como $\phi(x)$ é convexa então $H(\cdot|\mu)$ também é. Pelo fato de $H(\cdot|\mu)$ ser convexa decorre que esta é contínua em ν.

(ii) Pela desigualdade de Jensen, $\sum_x \phi(x) \geq \phi\left(\sum_x x\right) = \left(\sum_x x\right) \ln\left(\sum_x x\right)$, uma vez que $\phi(x)$ é convexa. Para $x = \frac{\nu(i)}{\mu(i)}$,

$$\sum_{i=1}^{r} \phi\left(\frac{\nu(i)}{\mu(i)}\right) \geq \left(\sum_{i=1}^{r} \frac{\nu(i)}{\mu(i)}\right) \ln\left(\sum_{i=1}^{r} \frac{\nu(i)}{\mu(i)}\right) > 0.$$

Justificando a última desigualdade, $\sum_{i=1}^{r} \frac{\nu(i)}{\mu(i)} > 0$ e

$$\ln\left(\sum_{i=1}^{r} \frac{\nu(i)}{\mu(i)}\right) \geq \ln\left(\frac{\nu(1)}{m} + \frac{\nu(2)}{m} + \cdots + \frac{\nu(r)}{m}\right) = \ln\left(\frac{1}{m}\right) > 0,$$

onde $m = \max_i \{\mu(i)\} > 0$ e $\sum_{i=1}^{r} \nu(i) = 1$. Portanto, $H(\cdot|\mu)$ é não negativa.
(iii) $H(\cdot|\mu)$ é finita, pois $\frac{\nu(i)}{\mu(i)}$ é finito se $\mu(i) > 0$, e também se $\mu(i) = 0$, uma vez que $0 \ln\left(\frac{0}{0}\right) \equiv 0$. □

3 Princípio de Grandes Desvios para Somas de Nível 1

Antes de estabelecermos um princípio de Grandes Desvios para somas de nível 1 em Cadeias de Markov será necessário enunciar um Teorema que fornece um Princípio de Grandes Desvios para seqüências de v.a's com qualquer tipo de interdependência.

Definição 3.1 *Seja $\{W_n\}_{n\geq 1}$ uma seqüência de v.a's, assumindo valores em $\mathbf{R}$, e definidas em $\{(\Omega_n, \mathcal{F}_n, \mathcal{P}_n)\}_{n\geq 1}$, sendo $(\Omega_n, \mathcal{F}_n, \mathcal{P}_n)$ um espaço de probabilidade. Para $a_n > 0$, com $\lim_{n\to\infty} a_n = \infty$, definimos*

$$c_n(t) = \frac{1}{a_n} \ln E_n\left(e^{tW_n}\right), \ t \in \mathbf{R}, \ \text{ onde } E_n\left(e^{tW_n}\right) = \int_{\Omega_n} e^{tW_n} dP_n\,.$$

Teorema 3.1 *Suponha que $c_n(t)$ existe e é finito para $\forall\, t \in \mathbf{R}$. Além disso, suponha que $c(t) = \lim_{n\to\infty} c_n(t)$ existe e é finito para $\forall\, t \in \mathbf{R}$. Então, $c_n(t)$ e $c(t)$ são funções convexas.*
Prova Segue da aplicação da desigualdade de Hölder, com $p = \frac{1}{1-\alpha}$ e $q = \frac{1}{\alpha}$, $\alpha \in (0,1)$. □

Definição 3.2 *A transformada de Legendre, ver Definição 2.14, para c(t) é $I(z) = \sup_{t\in\mathbf{R}} \{tz - c(t)\}$, $z \in \mathbf{R}$.*

Proposição 3.1 *Suponha que* $c(t)$ *existe e é finito,* $\forall t \in \mathbf{R}$. *Então*:
(i) I é convexa fechada;
(ii) $I(z) \geq 0$, $\forall z \in \mathbf{R}$;
(iii) $I(z) = 0$ se e somente se $z \in \partial c(0)$;
(iv) I é não crescente em $(-\infty, c'_{-}(0)]$ e não decrescente em $[c'_{+}(0), +\infty)$.

Prova

(i) Pelo Teorema 3.1 $c(t)$ é convexa e pela hipótese de $c(t)$ ser finita, tem-se que $c(t)$ é convexa fechada. Assim, pelo Teorema 2.4(i), segue que I é convexa fechada.
(ii) Para $\forall z \in \mathbf{R}$, quando $t = 0$ tem-se $0z - \ln c(0) = 0$, e portanto, $I(z) = \sup_{t \in \mathbf{R}} \{tz - c(t)\} \geq 0$.
(iii) Suponha que $I(z) = 0$, $z \in \mathbf{R}$. Assim, $\partial I(z) = \{\alpha \in \mathbf{R} : I(y) \geq 0 + \alpha(y - z), \forall y \in \mathbf{R}\}$. Mas $\alpha = 0$ é tal que $\alpha \in \partial I(z)$. Então, pelo Teorema 2.4 (ii), $z \in \partial c(0)$. Recíprocamente, suponha que $z \in \partial c(0)$. Então, pelo Teorema 2.4 (iii), $z0 = c(0) + I(z)$, ou seja, $I(z) = 0$.
(iv) Pela Proposição 2.2, $\partial c(0) = [c'_{-}(0), c'_{+}(0)]$. Pelo item (iii), $I(z) = 0$ para $z \in \partial c(0)$. Consideremos $c'_{+}(0) < z_1 < z_2$, com $z_1 = (1 - \lambda)c'_{+}(0) + \lambda z_2$, $\lambda \in (0, 1)$. Então, por (i)

$$I(z_1) = I\Big((1-\lambda)c'_{+}(0) + \lambda z_2\Big) \leq (1-\lambda)I(c'_{+}(0)) + \lambda I(z_2) = \lambda I(z_2) \leq I(z_2),$$

ou seja, I é não decrescente. Agora consideremos $z_1 < z_2 < c'_{-}(0)$, com $z_2 = (1 - \lambda)c'_{-}(0) + \lambda z_1$, $\lambda \in (0, 1)$. Então, por (i)

$$I(z_2) = I\Big((1-\lambda)c'_{-}(0) + \lambda z_1\Big) \leq (1-\lambda)I(c'_{-}(0)) + \lambda I(z_1) = \lambda I(z_1) \leq I(z_1),$$

ou seja, I é não crescente. □

Proposição 3.2 *A função* I*, da Definição 3.2, é uma função taxa, conforme a Definição 2.10.*

Prova

(i) I não é identicamente $+\infty$, pois pela Proposição 3.1 (iii), $I(z) = 0$ para $z \in \partial c(0)$.
(ii) I é semicontínua inferiormente. De fato, considere $\{y_n\}_{n \geq 1}$ uma seqüência em $\mathbf{R}$ tal que $\lim_{n \to \infty} y_n = y$, $y \in \mathbf{R}$. Então, para $\forall\, x \in \mathbf{R}$, $I(y_n) \geq xy_n - f(x)$ e $\liminf_{n \to \infty} I(y_n) \geq xy - f(x)$. Como x é arbitrário, $\liminf_{n \to \infty} I(y_n) \geq I(y)$, isto é, I é semicontínua inferiormente.
(iii) $K_L = \{z \in \mathbf{R} : I(z) \leq L\}$, $L \in (0, \infty)$, é fechado se para $\{y_n\}_{n \geq 1} \in K_L$, com $\lim_{n \to \infty} y_n = y$, tem-se que $y \in K_L$. Por (ii),

$L \geq \liminf_{n\to\infty} I(y_n) \geq I(y)$, ou seja, $y \in K_L$. Agora considere $z \in K_L$. Para $t = 1$ temos que $z - c(1) \leq L$ e para $t = -1$, $-z - c(1) \leq L$. Segue que $K_L \subset \{-L - c(-1), L + c(1)\}$, e portanto, K_L é limitado. Assim, concluímos que K_L é compacto. $\square$

Teorema 3.2 *Seja* $\{W_n\}_{n\to\infty}$ *uma seqüência de v.a's, assumindo valores em* $\mathbf{R}$*, e definidas em* $\{(\Omega_n, \mathcal{F}_n, \mathcal{P}_n)\}_{n\leq 1}$*. Suponha que* $c_n(t)$ *existe e é finito,* $\forall t \in \mathbf{R}$*, e também que* $c(t) = \lim_{n\to\infty} c_n(t)$ *existe e é finito para* $\forall t \in \mathbf{R}$*. Defina* $Q_n(A) = P_n\left(\frac{W_n}{a_n} \in A\right)$*, com* $A \in \mathcal{B}(\mathbf{R})$*, sendo* $\mathcal{B}(\mathbf{R})$ *a* $\sigma-$*álgebra de Borel em* $\mathbf{R}$*. Então*:
(i) $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\inf_{z\in K} I(z)$ para cada fechado $K \subset \mathbf{R}$;

(ii) assumindo que $c(t)$ é diferenciável, $\forall t \in \mathbf{R}$, $\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq -\inf_{z\in G} I(z)$ para cada aberto $G \subset \mathbf{R}$.
Prova
(i) Vamos denotar $[\alpha, \beta] = \partial c(0)$, isto é, $\alpha = c'_-(0)$ e $\beta = c'_+(0)$. A prova será dividida em 4 casos.

(1º) caso: $K = [v, +\infty)$, $v > \beta$.
Seja $t > 0$. Então, $Q_n(K) = P_n\left(e^{tW_n} \geq e^{ta_n v}\right) \leq e^{-a_n tv} E_n\left(e^{tW_n}\right) = e^{-a_n(tv - c_n(t))}$, sendo a desigualdade justificada pela Desigualdade de Chebyshev. Aplicando logarítmo e multiplicando $\frac{1}{a_n}$ na desigualdade, vem que $\frac{1}{a_n} \ln Q_n(K) \leq -tv + c_n(t)$. Segue que $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -tv + c(t)$. Como $t > 0$ é arbitrário,

$$\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq \inf_{t>0} \{-(tv - c(t))\} = -\sup_{t>0} \{tv - c(t)\}.$$

Uma vez que $\ln Q_n(K) \leq 0$ e quando $t = 0$ temos $-tv + c(t) = 0$, então $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\sup_{t\geq 0} \{tv - c(t)\}$. Como $\beta \in \partial c(0)$, decorre da Definição 2.13 que $c(t) \geq c(0) + \beta(t - 0) = \beta t$, e daí vem que $tv - c(t) < 0$ se $t < 0$. Assim, o supremo de $tv - c(t)$ está em $t \geq 0$, isto é, $\sup_{t\geq 0} \{tv - c(t)\} = \sup_{t\in\mathbf{R}} \{tv - c(t)\} = I(v)$. Outro fato é que $I(v) = \inf_{z\geq v} I(z) = \inf_{z\in K} I(z)$, decorrente da Proposição 3.1(iv). Dessas considerações, concluímos que $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\inf_{z\in K} I(z)$.

(2º) caso: $K = (-\infty, u]$, $u < \alpha$. O procedimento é análogo ao caso anterior.

(3º) caso: $K \cap [\alpha, \beta] \neq \emptyset$. Para $z \in [\alpha, \beta]$, sabe-se que $I(z) = 0$. Também sabemos que $\ln Q_n(K) \leq 0$. Logo, $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq 0 = -\inf_{z\in K} I(z)$.

(4º) caso: $K \cap [\alpha, \beta] = \emptyset$. Defina $u = \sup\{K \cap (-\infty, \alpha)\}$, com $u = -\infty$ se $K \cap (-\infty, \alpha) = \emptyset$, e $v = \inf\{K \cap (\beta, +\infty)\}$, com $v = +\infty$ se $K \cap (\beta, +\infty) = \emptyset$. Então, $K \subset (-\infty, u] \cup [v, +\infty)$, onde $u < \alpha \leq \beta < v$. Já sabemos, pelos casos (1º) e (2º), que $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n((-\infty, u]) \leq -I(u)$ e $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n([v, +\infty)) \leq -I(v)$. Agora, tem-se que provar:

$$\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\min\{I(u), I(v)\}.$$

Sem perda de generalidade suponha $I(u) < I(v)$. Seja $\epsilon > 0$. Pelo fato de

$$\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n((-\infty, u]) \leq -I(u) \quad \text{e} \quad \limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n([v, +\infty)) \leq -I(v),$$

então $\exists n_0 \in \mathbf{N}$ tal que para $n \geq n_0$, $\frac{1}{a_n} \ln Q_n((-\infty, u]) \leq -I(u) + \epsilon$ e $\frac{1}{a_n} \ln Q_n([v, +\infty)) \leq -I(v) + \epsilon$. Segue que $Q_n((-\infty, u]) \leq e^{-a_n(I(u)-\epsilon)}$ e $Q_n([v, +\infty)) \leq e^{-a_n(I(v)-\epsilon)} \leq e^{-a_n(I(u)-\epsilon)}$. Então, $Q_n(K) \leq 2e^{-a_n(I(u)-\epsilon)}$ para $n \geq n_0$, isto é, $\frac{1}{a_n} \ln Q_n(K) \leq \frac{\ln 2}{a_n} - I(u) + \epsilon$. Agora tome $n \geq m_0 \geq n_0$ tal que $a_n \geq \frac{\ln 2}{\epsilon}$. Daí vem que $\frac{1}{a_n} \ln Q_n(K) \leq -I(u) + 2\epsilon$. Como ϵ é arbitrário, $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -I(u)$, isto é,

$$\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\min\{I(u), I(v)\} = -\inf_{z\in K} I(z),$$

pois se $z < u$ temos $I(z) \geq I(u)$ e se $z > v$, $I(z) \geq I(v)$.

(ii) Para $t \in \mathbf{R}$ e $n \in \mathbf{N}^*$ defina $Q_{n,t}(A) = \int_A e^{a_n(tx - c_n(t))} dQ_n(x)$, onde $A \in \mathcal{B}(\mathbf{R})$. Seja $z \in G$, G aberto, e $\epsilon > 0$ tal que $B(z, \epsilon) = (z - \epsilon, z + \epsilon) \subset G$. Tome $x \in B(z, \epsilon)$. Então, $tx < tz + |t|\epsilon$, $\forall t \neq 0$.

Segue que

$$
\begin{aligned}
Q_n(G) &\geq Q_n(B(z,\epsilon)) = \int_{B(z,\epsilon)} dQ_n(x) = \int_{B(z,\epsilon)} e^{-a_n(tx-c_n(t))} dQ_{n,t}(x) \\
&= e^{a_n c_n(t)} \int_{B(z,\epsilon)} e^{a_n(-tx)} dQ_{n,t}(x) \\
&> e^{a_n c_n(t)} \int_{B(z,\epsilon)} e^{a_n(-tz-|t|\epsilon)} dQ_{n,t}(x) \\
&= e^{a_n c_n(t) - a_n(tz+|t|\epsilon)} \int_{B(z,\epsilon)} dQ_{n,t}(x) \\
&= e^{a_n c_n(t) - a_n(tz+|t|\epsilon)} Q_{n,t}(B(z,\epsilon)).
\end{aligned}
$$

Aplicando logarítmo e multiplicando por $\frac{1}{a_n}$, tem-se $\frac{1}{a_n} \ln Q_n(G) > c_n(t) - tz - |t|\epsilon + \frac{1}{a_n} \ln Q_{n,t}((B(z,\epsilon))$, e portanto,

$$
\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) > \liminf_{n\to\infty} c_n(t) - tz - |t|\epsilon + \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_{n,t}((B(z,\epsilon)).
$$

Por hipótese temos que $\lim_{n\to\infty} c_n(t) = c(t)$. Agora, pelo fato de supormos $c(t)$ derivável para $\forall\, t \in \mathbf{R}$, pela Proposição 2.3(iii) segue que $\exists\, \tau \in \mathbf{R}$ tal que $z = c^{'}(\tau)$. Assim, $I(z) = z\tau - c(\tau)$. Logo,

$$
\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq -I(z) + \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_{n,t}(B(z,\epsilon)),
$$

uma vez que ϵ é arbitrário. Também pelo fato de $z = c^{'}(\tau)$, $\lim_{n\to\infty} Q_{n,t}(B(z,\epsilon)) = 1$, cuja prova pode ser vista em Ellis (1985). Como conseqüência $\lim_{n\to\infty} \frac{1}{a_n} \ln Q_{n,t}(B(z,\epsilon)) = 0$. Prosseguindo, tem-se que $\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq -I(z)$. Como $z \in G$ é arbitrário,

$$
\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq \sup_{z\in G} \{-I(z)\} = -\inf_{z\in G} I(z)\,.
$$

Comentário 3.1 Seja $A \subset \mathbf{R}$. Uma vez que $int(A)$ é aberto temos que

$$
-\inf_{z\in int(A)} I(z) \leq \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(int(A))\,,
$$

e por $\bar{A}$ ser fechado $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(\bar{A}) \leq -\inf_{z\in\bar{A}} I(z)$. Então:

$$\begin{aligned} -\inf_{z\in int(A)} I(z) &\leq \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(int(A)) \leq \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(A) \\ &\leq \limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(A) \leq \limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(\bar{A}) \\ &\leq -\inf_{z\in\bar{A}} I(z)\,. \end{aligned}$$

Portanto, o limite existe se $\inf_{z\in int(A)} I(z) = \inf_{z\in\bar{A}} I(z)$.

Comentário 3.2 Suponha $A = (-\infty, z]$, com $z \leq c'(0)$. Então, $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(A) \leq -\inf_{y\in A} I(y) = -I(z)$, onde a última igualdade justifica-se pela Proposição 3.1 (iv). Seja $\delta > 0$. Segue que

$$\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n((-\infty, z+\delta]) \geq \liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n((z-\delta, z+\delta)) \geq -\inf_{y\in(z-\delta,z+\delta)} I(y).$$

Como δ é arbitrário, $\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n((-\infty, z]) \geq -I(z)$, e assim concluímos que $\lim_{n\to\infty} \frac{1}{a_n} \ln Q_n(A) = -I(z)$, ou seja, o limite existe para intervalos fechados.

Comentário 3.3 Considere $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. definidas em $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathcal{P})$, com $E|X_1| < \infty$, $a_n = n$ e $W_n = X_1 + \cdots + X_n$. Assim, $c_n(t) = \ln E\left(e^{tX_1}\right) = c(t)$, $\forall n \geq 1$. Então, $\frac{d}{dt}c(t)\big|_{t=0} = EX_1$. Para $x \leq EX_1$ temos, pelo Comentário 3.2, que $\lim_{n\to\infty} \frac{1}{n} \ln P\left(\frac{W_n}{n} \leq x\right) = -I(x) = -\sup_{t\in\mathbf{R}} \{xt - c(t)\}$, ou seja, o mesmo resultado do Teorema de Cramer-Chernoff, conforme Dembo & Zeitouni (1993).

Teorema 3.3 *Seja $\{X_n\}_{n\geq 0}$ uma Cadeia de Markov irredutível e aperiódica, com espaço de estados $S = \{1, \cdots, r\}$ e matriz de transição $P = \{p(i,j)\}_{i,j\in S}$. Defina $S_n = \sum_{i=1}^n X_i$, $P_t = \{p_t(i,j)\}_{i,j\in S}$, com $p_t(i,j) = e^{tj}p(i,j)$, $t \in \mathbf{R}$, e $c_{n,\sigma}(t) = \frac{1}{n} \ln E_\sigma\left(e^{tS_n}\right)$, onde $\sigma \in S$ é o estado inicial que a cadeia assume. A seqüência $\left\{\bar{X}_n\right\}_{n\geq 1}$, onde $\bar{X}_n = \frac{S_n}{n}$, satisfaz o PGD com função taxa $I(z) = \sup_{t\in\mathbf{R}} \{tz - \ln\lambda(P_t)\}$, $z \in \mathbf{R}$, sendo $\lambda(P_t)$ o autovalor de Perron-Frobenius, isto é*:
(i) $\limsup_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in F) \leq -\inf_{x\in F} I(x)$, $F \subset \mathbf{R}$ fechado;
(ii) $\liminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in G) \geq -\inf_{x\in G} I(x)$, $G \subset \mathbf{R}$ aberto.

Prova Pela Definição 3.2, $I(z) = \sup_{t\in\mathbf{R}} \{tz - c(t)\}$. Assim, temos que provar que $\lim_{n\to\infty} c_{n,\sigma}(t) = \ln \lambda(P_t)$. De fato,

$$\begin{aligned}
c_{n,\sigma}(t) &= \frac{1}{n}\ln\left(\sum_{(x_1,\cdots,x_n)\in S} P_\sigma(X_1=x_1,\cdots,X_n=x_n)e^{t\sum_{k=1}^n x_k}\right) \\
&= \frac{1}{n}\ln\left(\sum_{(x_1,\cdots,x_n)\in S} P_\sigma(X_1=x_1,\cdots,X_n=x_n)\prod_{k=1}^n e^{tx_k}\right) \\
&= \frac{1}{n}\ln\left(\sum_{x_1\in S}\cdots\sum_{x_n\in S} p(\sigma,x_1)e^{tx_1}\cdots p(x_{n-1},x_n)e^{tx_n}\right) \\
&= \frac{1}{n}\ln\left(\sum_{x_n\in S} p_t^n(\sigma,x_n)\right).
\end{aligned}$$

Pelo Teorema 2.1(v) aplicado à matriz P_t, com $\phi = (1,\cdots,1)$, segue que

$$\lim_{n\to\infty} c_{n,\sigma}(t) = \lim_{n\to\infty}\frac{1}{n}\ln\left(\sum_{x_n\in S} p_t^n(\sigma,x_n)\right) = \ln\lambda(P_t).$$

Pode-se provar, ver Lancaster (1969), que $\lambda(P_t)$ é finito e diferenciável com relação a t. Prosseguindo, pelo Teorema 3.2, temos que (i) e (ii) valem com função taxa $I(z) = \sup_{t\in\mathbf{R}}\{tz - \ln\lambda(P_t)\}$. □

Proposição 3.3 *Seja $\{X_n\}_{n\geq 0}$ uma Cadeia de Markov finita, irredutível e aperiódica. Para $z = \sum_{x\in S} x\mu(x)$, tem-se $I(z) = 0$, sendo μ a medida invariante da Cadeia.*
Prova Pela desigualdade de Jensen, $c_{n,\sigma}(t) = \frac{1}{n}\ln E_\sigma\left(e^{tS_n}\right) \geq \frac{1}{n}E_\sigma(tS_n) = tE_\sigma\left(\frac{S_n}{n}\right)$. Mas

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n\sum_{x\in S} x\delta_{X_i}(x) = \sum_{x\in S} x\sum_{i=1}^n\frac{\delta_{X_i}(x)}{n} = \sum_{x\in S} xL_n(x),$$

sendo a última igualdade justificada pela Definição 2.5. Pelo Teorema 2.3(i), $\lim_{n\to\infty} E_\sigma(L_n(x)) = \mu(x)$, e portanto,

$$\lim_{n\to\infty} E_\sigma\left(\frac{S_n}{n}\right) = \sum_{x\in S} x\lim_{n\to\infty} E_\sigma(L_n(x)) = \sum_{x\in S} x\mu(x).$$

Segue que $tz - c_{n,\sigma}(t) \leq tz - tE_\sigma\left(\frac{S_n}{n}\right)$. Quando $n\to\infty$, $tz - \ln\lambda(P_t) \leq t(z - \sum_{x\in S} x\mu(x)) = 0$, ou seja, $tz - \ln\lambda(P_t) \leq 0$ para $\forall t\in\mathbf{R}$. Assim, $I(z) = 0$. □

4 Princípio de Grandes Desvios para Somas de Nível 2

Para estabelecermos um PGD para somas de nível 2 em Cadeias de Markov será preciso um resultado análogo ao Teorema 3.2, mas com $\{W_n\}_{n\geq 1}$ definida em $\left\{(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \mathcal{P}_n)\right\}_{n\geq 1}$.

Definição 4.1 *Seja $\{W_n\}_{n\geq 1}$ uma seqüência de v.a's, assumindo valores em $\mathbf{R}^d$, e definidas no espaço de probabilidade $\{(\Omega_n, \mathcal{F}_n, \mathcal{P}_n)\}_{n\geq 1}$. Para $a_n > 0$, com $\lim_{n\to\infty} a_n = +\infty$, definimos*

$$c_n(t) = \frac{1}{a_n} \ln E_n\left(e^{\langle t, W_n\rangle}\right) \quad e \quad E_n\left(e^{\langle t, W_n\rangle}\right) = \int_{\Omega_n} e^{\langle t, W_n\rangle} dP_n\,, t\in \mathbf{R}^d$$

sendo $\langle t, W_n\rangle$ o produto interno usual em R^d.

Comentário 4.1 Com as hipóteses de que $c_n(t)$ e $c(t) = \lim_{n\to\infty} c_n(t)$ existem e são finitas para $\forall t \in \mathbf{R}^d$, prova-se, analogamente ao Teorema 3.1, que $c_n(t)$ e $c(t)$ são funções convexas.

Definição 4.2 *A transformada de Legendre para $c(t)$ é $I(z) = \sup_{t\in\mathbf{R}^d}\{\langle t, z\rangle - c(t)\}\ , z\in \mathbf{R}^d$.*

Comentário 4.2 Com as hipóteses de que $c_n(t)$ e $c(t)$ existem e são finitas para $\forall t \in \mathbf{R}^d$, também prova-se que I é uma função taxa, além de ser convexa fechada.

Teorema 4.1 *Seja $\{W_n\}_{n\geq 1}$ uma seqüência de v.a's, assumindo valores em $\mathbf{R}^d$, definidas em $\{(\Omega_n, \mathcal{F}_n, \mathcal{P}_n)\}_{n\geq 1}$. Suponha que $c_n(t)$ e $c(t) = \lim_{n\to\infty} c_n(t)$ existem e são finitas para todo $t\in\mathbf{R}^d$. Defina $Q_n(A) = P_n\left(\frac{W_n}{a_n}\in A\right)$, $A\in\mathcal{B}(\mathbf{R}^d)$. Então:*
(i) $\limsup_{n\to\infty} \frac{1}{a_n} \ln Q_n(K) \leq -\inf_{z\in K} I(z)$ para cada fechado $K\subset \mathbf{R}^d$;
(ii) assumindo que $c(t)$ é diferenciável, $\forall t\in\mathbf{R}^d$,

$$\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq -\inf_{z\in G} I(z),$$

para cada aberto $G\subset\mathbf{R}^d$.
Prova (i) Na prova do Teorema 3.2, primeiro foram considerados conjuntos fechados do tipo $(-\infty, x]$ e $[x, +\infty)$, e depois extendeu-se

para fechados quaisquer. No caso $d > 1$, as semi-retas são trocadas por certos semi-espaços fechados.

Denotemos $I(K) \equiv \inf_{z\in K} I(z)$. Para $I(K) = 0$ concluímos imediatamente que a desigualdade vale. Por isso, assumiremos $I(K) > 0$. Prosseguindo com a prova utilizaremos o seguinte resultado, conforme Ellis (1985):

Seja K um fechado não vazio em $\mathbf{R}^d$. Dado $t \neq (0, \cdots, 0) \in \mathbf{R}^d$ e um real α defina o semi-espaço fechado $H(t,\alpha) = \left\{z \in \mathbf{R}^d : \langle t, z\rangle - c(t) \geq \alpha\right\}$. Assim, tem-se:

(1º) se $0 < I(K) < \infty$ então para qualquer $0 < \epsilon < I(K)$ existem $t_1, \cdots, t_r \in \mathbf{R}^d$, não nulos, tais que $K \subseteq \bigcup_{i=1}^r H(t_i, I(K) - \epsilon)$.

(2º) se $I(K) = +\infty$ então para qualquer $R > 0$ existem $t_1, \cdots, t_r \in \mathbf{R}^d$, não nulos, tais que $K \subseteq \bigcup_{i=1}^r H(t_i, R)$.

Agora assuma que $0 < I(K) < +\infty$. Portanto,

$$\begin{aligned} Q_n(K) &\leq \sum_{i=1}^r P_n\left(\frac{W_n}{a_n} \in H(t_i, I(K) - \epsilon)\right) \\ &= \sum_{i=1}^r P_n\left(\langle t_i, W_n\rangle \geq a_n(c(t_i) + I(K) - \epsilon)\right) \\ &\leq \sum_{i=1}^r e^{-a_n(c(t_i)+I(K)-\epsilon)} E_n\left(e^{\langle t_i, W_n\rangle}\right) \\ &= \sum_{i=1}^r e^{a_n(c_n(t_i)-c(t_i)-I(K)+\epsilon)} = e^{-a_n I(K)} \sum_{i=1}^r e^{a_n(c_n(t_i)-c(t_i)+\epsilon)}, \end{aligned}$$

sendo a última desigualdade justificada pela desigualdade de Chebyshev. Para $j \in \{1, \cdots, r\}$ tal que $c_n(t_j) - c(t_j) \geq c_n(t_i) - c(t_i), \forall i \in \{1, \cdots, r\}$, temos $Q_n(K) \leq e^{-a_n I(K)} r e^{a_n(c_n(t_j)-c(t_j)+\epsilon)}$. Aplicando logarítmo e dividindo por a_n, $\frac{1}{a_n} \ln Q_n(K) \leq \frac{1}{a_n} \ln r + (c_n(t_j) - c(t_j) + \epsilon) - I(K)$. Como $\lim_{n\to\infty} c_n(t_j) = c(t_j)$ e $\epsilon \in (0, I(K))$ é arbitrário, segue o resultado.

Para $I(K) = +\infty$ o procedimento é análogo, utilizando (2º) e o fato de que $R > 0$ é arbitrário.

(ii) Suponha que $c(t)$ é diferenciável para $\forall\, t \in \mathbf{R}^d$. Analogamente ao Teorema 3.2, define-se $Q_{n,t}(dx) = e^{a_n\langle t,x\rangle}\, Q_n(dx) e^{-a_n c_n(t)}$, $t \in \mathbf{R}^d$ e $n \geq 1$. Também definimos $B(z,\epsilon)$ como uma bola aberta de raio $\epsilon > 0$ centrada em $z = \bigtriangledown c(t)$, sendo $\bigtriangledown c(t)$ é o gradiente de $c(t)$. Prova-se, ver Ellis (1985), que $\lim_{n\to\infty} Q_{n,t}(B(z,\epsilon)) = 1$ e, utilizando este resultado, chega-se a $\liminf_{n\to\infty} \frac{1}{a_n} \ln Q_n(G) \geq -\inf_{z\in G} I(z)$. $\square$

Teorema 4.2 *Seja* $\{X_n\}_{n\geq 0}$ *uma Cadeia de Markov irredutível, aperiódica, com espaço de estados* $S = \{1, \cdots, r\}$ *e matriz de transição* $P = \{p(i,j)\}_{i,j\in S}$. *Defina* $S_n = \sum_{i=1}^{n} f(X_i)$, *onde*

$$f(X_i) = (\delta_{X_i}(1), \cdots, \delta_{X_i}(r)) \quad e \quad \delta_{X_i}(j) = \begin{cases} 1, & se\ X_i = j \\ 0, & caso\ contrário \end{cases}, j \in \{1, \cdots, r\}.$$

Também defina $P_t = \{p_t(i,j)\}_{i,j\in S}$, $p_t(i,j) = e^{\langle t, f(j)\rangle} p(i,j)$, *e* $c_{n,\sigma}(t) = \frac{1}{n} \ln E_\sigma \left(e^{\langle t, S_n\rangle}\right)$, *com* $t \in \mathbf{R}^r$ *e* $\sigma \in S$ *sendo o estado incial que a Cadeia assume. Então a seqüência* $\left\{\bar{X}_n\right\}_{n\geq 1}$, $\bar{X}_n = \frac{S_n}{n}$, *satisfaz o PGD com função taxa* $I(q) = \sup_{t\in\mathbf{R}^r} \{\langle t, q\rangle - \ln \lambda(P_t)\}$, *sendo* $q \in \mathcal{M}(\mathcal{S}) \equiv \{$ *espaço de todas medidas de probabilidade em* S $\}$ *e* $\lambda(P_t)$ *o autovalor de Perron-Frobenius, isto é:*

(i) $\liminf_{n\to\infty} \frac{1}{n} \ln P\left(\frac{S_n}{n} \in F\right) \leq -\inf_{q\in F} I(q)$ para F fechado tal que $F \in \mathcal{B}(\mathcal{M}(\mathcal{S}))$;

(ii) $\limsup_{n\to\infty} \frac{1}{n} \ln P\left(\frac{S_n}{n} \in G\right) \geq -\inf_{q\in G} I(q)$ para G aberto tal que $G \in \mathcal{B}(\mathcal{M}(\mathcal{S}))$.

Prova Observe que $f : S \longrightarrow \mathbf{R}^r$. Logo, $\left\{\bar{X}_n\right\}_{n\geq 1}$ é uma seqüência de v.a's assumindo valores em $\mathbf{R}^r$. Pela definição 4.2, $I(q) = \sup_{t\in\mathbf{R}^r} \{\langle t, q\rangle - c(t)\}$, $q \in \mathcal{M}(\mathcal{S})$. Portanto, temos que provar $\lim_{n\to\infty} c_{n,\sigma}(t) = \ln \lambda(P_t)$. Mas

$$\begin{aligned} c_{n,\sigma}(t) &= \frac{1}{n} \ln \left[\sum_{(x_1,\cdots,x_n)\in S} P_\sigma(X_1 = x_1, \cdots, X_n = x_n) \prod_{k=1}^{n} e^{\langle t, f(x_k)\rangle} \right] \\ &= \frac{1}{n} \ln \left[\sum_{x_1\in S} \cdots \sum_{x_n\in S} p(\sigma, x_1) e^{\langle t, f(x_1)\rangle} \cdots p(x_{n-1}, x_n) e^{\langle t, f(x_n)\rangle} \right] \\ &= \frac{1}{n} \ln \left[\sum_{x_n\in S} p_t^n(\sigma, x_n) \right]. \end{aligned}$$

Aplicando o Teorema 2.1(v) a P_t, com $\phi = (1, \cdots, 1)$, temos $\lim_{n\to\infty} c_{n,\sigma}(t) = \ln \lambda(P_t)$. Como $\lambda(P_t)$ é finito e diferenciável com relação a t, ver Lancaster (1969), temos que valem (i) e (ii) do Teorema 4.1, com função taxa $I(q) = \sup_{t\in\mathbf{R}^r} \{\langle t, q\rangle - \ln \lambda(P_t)\}$, $q \in \mathcal{M}(\mathcal{S})$. $\square$

Proposição 4.1 *Seja* $\{X_n\}_{n\geq 0}$ *uma C.M. finita, irredutível e aperiódica. Para* $q = \mu$, *sendo* $\mu = (\mu(1), \cdots, \mu(r))$ *a medida invariante da Cadeia, tem-se* $I(q) = 0$.

Prova Note que $S_n = \left(\sum_{i=1}^n \delta_{X_i}(1), \cdots, \sum_{i=1}^n \delta_{X_i}(r)\right)$ e $\frac{S_n}{n} = \left(\frac{\sum_{i=1}^n \delta_{X_i}(1)}{n}, \cdots, \frac{\sum_{i=1}^n \delta_{X_i}(r)}{n}\right)$. Pela desigualdade de Jensen,

$$\begin{aligned} c_{n,\sigma}(t) &= \frac{1}{n} \ln E_\sigma \left(e^{\langle t, S_n\rangle}\right) \geq \frac{1}{n} E_\sigma \left(\ln e^{\langle t, S_n\rangle}\right) \\ &= \frac{1}{n} E_\sigma \left(\langle t, S_n\rangle\right) = E_\sigma \left(\langle t, \frac{S_n}{n}\rangle\right) \\ &= \left(t_1 E_\sigma \left(\frac{\sum_{i=1}^n \delta_{X_i}(1)}{n}\right), \cdots, t_r E_\sigma \left(\frac{\sum_{i=1}^n \delta_{X_i}(r)}{n}\right)\right), \forall t \in \mathbf{R}^r . \end{aligned}$$

Segue que $\langle t, q\rangle - c_{n,\sigma}(t) \leq \langle t, q\rangle - E_\sigma \left(\langle t, \frac{S_n}{n}\rangle\right)$. Já sabemos que $\lim_{n\to\infty} c_{n,\sigma}(t) = \ln \lambda(P_t)$. Agora, $\lim_{n\to\infty} E_\sigma \left(\frac{S_n}{n}\right) = \mu$, uma vez que $\lim_{n\to\infty} E_\sigma \left(\frac{\sum_{i=1}^n \delta_{X_i}(j)}{n}\right) = \mu(j)$, $j \in S$, conforme o Teorema 2.3 (i). Assim, $\langle t, q\rangle - \ln \lambda(P_t) \leq \langle t, q\rangle - \langle t, q\rangle = 0\, , \forall t \in \mathbf{R}^r$. Logo, $I(q) = 0$. $\square$

O teorema a seguir apresenta uma expressão alternativa para a função taxa I, que adiante será útil para estabelecermos o PGD para somas de nível 3.

Teorema 4.3 *A função taxa I, do Teorema 4.2, é tal que*

$$I(q) = \sup_{u \gg 0} \sum_{j=1}^r q_j \ln\left(\frac{u_j}{(uP)_j}\right) \qquad \textit{para} \quad q \in \mathcal{M}(\mathcal{S})\, ,$$

"$\gg$" denotando um vetor com entradas estritamente positivas e $(uP)_j = \sum_{i=1}^r u_i p(i,j)$.

Prova Seja $q \in \mathcal{M}(\mathcal{S})$ e $u \gg 0$. Pelo fato de $u \gg 0$ e P ser uma matriz irredutível e aperiódica, tem-se $uP \gg 0$. Defina a matriz $P_t = \{p_t(i,j)\}_{i,j\in S}$, com $p_t(i,j) = p(i,j)e^{t_j}$ e $t_j = \ln\left(\frac{u_j}{(uP)_j}\right)$. Segue que

$$uP_t = \left(\sum_{i=1}^r u_i p(i,1) \frac{u_1}{(uP)_1}, \cdots, \sum_{i=1}^r u_i p(i,r) \frac{u_1}{(uP)_r}\right) = (u_1, \cdots, u_r) = u\, .$$

Recursivamente: $uP_t^2 = uP_tP_t = uP_t = u$, $uP_t^3 = uP_t^2 P_t = uP_t = u, \cdots, uP_t^n = uP_t^{n-1}P_t = uP_t = u$. Assim,

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(\sum_{i=1}^r u_i p_t^n(i,j)\right) = \lim_{n\to\infty} \frac{1}{n} \ln u_j = 0$$

e, pelo Teorema 2.1(v), $\ln \lambda(P_t) = 0$, isto é, $\lambda(P_t) = 1$. Prosseguindo,

$$I(q) = \sup_{t\in\mathbf{R}^r} \{\langle t, q\rangle - \ln \lambda(P_t)\} = \sup_{t\in\mathbf{R}^r} \{\langle t, q\rangle\} = \sup_{t\in\mathbf{R}^r} \left\{\sum_{j=1}^r q_j t_j\right\} \geq \sum_{j=1}^r q_j \ln\left(\frac{u_j}{(uP)_j}\right) .$$

Mas uma vez que $u \gg 0$ é arbitrário, a última desigualdade implica que $I(q) \geq \sup_{u \gg 0} \sum_{j=1}^r \ln \left(\frac{u_j}{(uP)_j} \right)$.

Para estabelecer a desigualdade recíproca fixe um vetor $t \in \mathbf{R}^r$ e assuma $v \gg 0$ um autovetor à esquerda, correspondente ao autovalor $\lambda(P_t)$ da matriz irredutível P_t. Isto quer dizer que $vP_t = \lambda(P_t)v$. Agora observe que

$$\begin{aligned}
\langle t, q\rangle &+ \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j}{v_j} \right) = \sum_{j=1}^r t_i q_j + \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j}{v_j} \right) \\
&= \sum_{j=1}^r q_j \ln e^{t_j} + \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j}{v_j} \right) = \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j e^{t_j}}{v_j} \right) \\
&= \sum_{j=1}^r q_j \ln \left(\frac{(vP_t)_j}{v_j} \right) = \sum_{j=1}^r q_j \ln \left(\frac{\lambda(P_t) v_j}{v_j} \right) \\
&= \sum_{j=1}^r q_j \ln \lambda(P_t) = \ln \lambda(P_t) \,,
\end{aligned}$$

ou seja, $\ln \lambda(P_t) = \sum_{j=1}^r q_j \ln \left(\frac{(vP_t)_j}{v_j} \right)$. Então,

$$\begin{aligned}
\langle t, q\rangle - \ln \lambda(P_t) &= \langle t, q\rangle - \sum_{j=1}^r q_j \ln \left(\frac{(vP_t)_j}{v_j} \right) = \langle t, q\rangle - \left(\langle t, q\rangle + \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j}{v_j} \right) \right) \\
&= - \sum_{j=1}^r q_j \ln \left(\frac{(vP)_j}{v_j} \right) = \sum_{j=1}^r q_j \ln \left(\frac{v_j}{(vP)_j} \right) \,,
\end{aligned}$$

e portanto, $\langle t, q\rangle - \ln \lambda(P_t) \leq \sup_{u \gg 0} \sum_{j=1}^r q_j \ln \left(\frac{u_j}{(uP)_j} \right)$. □

Comentário 4.2 A expressão alternativa para I, no teorema acima, também é conhecida como "função entropia"e tem como interpretação o grau de afastamento de $q \in \mathcal{M}(\mathcal{S})$ com relação à medida invariante da cadeia.

5 Princípio de Grandes Desvios para Somas de Nível 3

No decorrer desta seção assuma $\{X_n\}_{n \geq 0}$ uma Cadeia de Markov com espaço de estados $S = \{1, \cdots, r\}$ e matriz de transição $P = \{p(i,j)\}_{i,j=1}^r$ tal que $p(i,j) > 0 \,, \forall i, j \in S$.

Denotemos

$$Z_i = \begin{cases} (X_{i-1}, X_i), & \text{para} \quad i = 1, \cdots, n-1 \\ (X_0, X_n), & \text{para} \quad i = n \end{cases} .$$

Então o espaço de estados de $\{Z_n\}_{n\geq 1}$ é S^2=$\{(1,1),(1,2),\cdots,(r,r)\}$. Já a matriz de transição de $\{Z_n\}_{n\geq 1}$ será denotada por $Q = \{q((k,l),(i,j))\}_{i,j,k,l=1}^{r}$.

Note que

$$
\begin{aligned}
q((k,l),(i,j)) &= P\left(Z_{n+1}=(i,j)|Z_n=(k,l)\right) \\
&= P\left(X_n=i, X_{n+1}=j|X_{n-1}=k, X_n=l\right) \\
&= \frac{P\left(X_{n+1}=j, X_n=i, X_n=l, X_{n-1}=k\right)}{P\left(X_{n-1}=k, X_n=l\right)} \\
&= \frac{P\left(X_{n+1}=j, X_n=i, X_{n-1}=k\right)}{P\left(X_{n-1}=k, X_n=l\right)} \\
&= P\left(X_{n+1}=j|X_n=i\right)=p(i,j) \quad \text{se} \quad i=l
\end{aligned}
$$

e zero caso contrário. Sendo assim,

$$
q((k,l),(i,j)) = \delta_l(i)\,p(i,j)\,, \quad \delta_l(i) = \begin{cases} 1, & \text{se} \quad i=l \\ 0, & \text{caso contrário} \end{cases} \tag{5.1}
$$

Para transições de $n \geq 2$ passos, tem-se

$$
\begin{aligned}
q^n\left((k,l),(i,j)\right) &= P\left(Z_{n+1}=(i,j)|Z_1=(k,l)\right)=P\left(X_{n+1}=j, X_n=i|X_0=k, X_1=l\right) \\
&= P\left(X_n=i|X_1=l\right)P\left(X_{n+1}=j|X_n=i\right)=p^{n-1}(l,i)p(i,j)\,.
\end{aligned}
$$

A Cadeia $\{Z_n\}_{n\geq 1}$ também é irredutível, pois para (k,l) e (i,j) $\exists m>0$ tal que $q^m((k,l),(i,j)) = p^{m-1}(l,i)p(i,j) > 0$, uma vez que todas entradas de P são positivas. A medida invariante ν de $\{Z_n\}_{n\geq 1}$ é $\nu(i,j)=p(i,j)\mu(i)$, pois

$$
\lim_{n\to\infty} q^n((k,l),(i,j)) = \lim_{n\to\infty} p^{n-1}(l,i)p(i,j) = p(i,j)\mu(i)\,,
$$

sendo μ a medida invariante de $\{X_n\}_{n\geq 0}$.

Definição 5.1 *Uma medida de probabilidade $\nu \in \mathcal{M}(\mathcal{S}^{\in}) = \{$ espaço de todas medidas de probabilidade em S^2 $\}$ é dita invariante por transição se $\sum_{k=1}^{r}\nu(k,i) = \sum_{k=1}^{r}\nu(i,k)$, $\forall i \in S$.*

Proposição 5.1 *A medida invariante ν da Cadeia $\{Z_n\}_{n\geq 1}$ é invariante por transição.*

Prova $\sum_{k=1}^{r}\nu(i,k) = \sum_{k=1}^{r}\mu(i)p(i,k) = \mu(i)\sum_{k=1}^{r}p(i,k) = \mu(i)$. Por outro lado, $\sum_{k=1}^{r}\nu(k,i) = \sum_{k=1}^{r}\mu(k)p(k,i) = \mu(i)$, uma vez que μ é a medida invariante de $\{X_n\}_{n\geq 0}$. $\square$

Exemplo 5.1 Suponha uma Cadeia de Markov $\{X_n\}_{n\geq 0}$ com $S = \{1,2\}$ e matriz de transição

$$P = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

A matriz de transição da Cadeia $\{Z_n\}_{n\geq 1}$ é

$$Q = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

A medida invariante de $\{X_n\}_{n\geq 0}$ é $\mu = \left(\frac{1}{2}, \frac{1}{2}\right)$ e de $\{Z_n\}_{n\geq 1}$ é $\nu = \left(\frac{2}{6}, \frac{1}{6}, \frac{1}{6}, \frac{2}{6}\right)$, ou seja, $\nu(1,1) = \frac{2}{6}$, $\nu(1,2) = \frac{1}{6}$, $\nu(2,1) = \frac{1}{6}$ e $\nu(2,2) = \frac{2}{6}$. $\square$

Quando a matriz de transição P tiver ao menos uma entrada nula então a Cadeia $\{Z_n\}_{n\geq 1}$ não será mais irredutível. O exemplo a seguir ilustra esta afirmação.

Exemplo 5.2 Considere uma Cadeia de Markov $\{X_n\}_{n\geq 0}$ com $S = \{1,2\}$ e matriz de transição

$$P = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ 1 & 0 \end{bmatrix}.$$

A matriz de transição de $\{Z_n\}_{n\geq 1}$ é

$$Q = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Claramente $\{Z_n\}_{n\geq 1}$ não é irredutível, mas é ergódica e sua medida invariante é $\nu = (0,14 \quad 0,43 \quad 0,43 \quad 0)$. $\square$

O próximo objetivo é estabelecer um PGD para a seqüência $\left\{\bar{Z}_n\right\}_{n\geq 1}$, sendo $\bar{Z}_n = \frac{S_n}{n}$,

$$S_n = \sum_{k=1}^{n} \left(\delta_{Z_k(1,1)}, \delta_{Z_k(1,2)}, \cdots, \delta_{Z_k(r,r)}\right) \text{ e } \delta_{Z_k}(i,j) = \begin{cases} 1, & \text{se } Z_k = (i,j) \\ 0, & \text{caso contrário} \end{cases}.$$

Comentário 5.1 Note que a seqüência $\{Z_n\}_{n\geq 1}$ enquadra-se no Teorema 4.2, uma vez que esta pode ser vista como uma Cadeia irredutível, aperiódica, com espaço de estados $T = \{1, 2, \cdots, s\}$, $s = r^2$, e matriz de transição $Q = \{q(i,j)\}_{i,j=1}^{s}$. Sendo assim, $\left\{\bar{Z}_n\right\}_{n\geq 1}$ satisfaz (i) e (ii) do Teorema 4.2, para $F \in \mathcal{B}\left(\mathcal{M}(\mathcal{S}^{\in})\right)$, $G \in \mathcal{B}\left(\mathcal{M}(\mathcal{S}^{\in})\right)$ e função taxa

$$I(\nu) = \sup_{t\in\mathbf{R}^s} \left\{\langle t, \nu\rangle - \ln \lambda(Q_t)\right\} \text{ , com } \nu \in \mathcal{M}(\mathcal{T})\,.$$

Além disso, pela Proposição 4.1, $I(\nu) = 0$ se ν for a medida invariante de $\{Z_n\}_{n\geq 1}$, a saber $\nu(i,j) = \mu(i)p(i,j)$.

Analogamente à função taxa para somas de nível 2, para somas de nível 3 existe uma expressão alternativa para $I(\nu)$, em termos de entropia relativa.

Teorema 5.1 *Seja $\{X_n\}_{n\geq 0}$ uma Cadeia de Markov com espaço de estados $S = \{1, \cdots, r\}$ e matriz de transição $P = \{p(i,j)\}_{i,j=1}^{r}$ tal que $p(i,j) > 0\,, \forall\, i,j \in S$. Defina a Cadeia $\{Z_n\}_{n\geq 1}$, com $Z_n = (X_{n-1}, X_n)$, onde o espaço de estados é $S^2 = \{(1,1), \cdots, (r,r)\}$ e a matriz de transição Q é dada por (5.1). Então, para toda medida de probabilidade $\nu \in \mathcal{M}(\mathcal{S}^{\in})$, tem-se*

$$I(\nu) = \begin{cases} \sum_{i=1}^{r} \nu_c(i) \sum_{j=1}^{r} \frac{\nu(i,j)}{\nu_c(i)} \ln\left(\frac{\nu(i,j)}{(\nu P)_{i,j}}\right), & \text{se } \nu \text{ for invariante por transição} \\ +\infty, & \text{caso contrário} \end{cases},$$

sendo $\nu_c(i) = \sum_{k=1}^{r} \nu(i,k)$, $\nu_l(i) = \sum_{k=1}^{r} \nu(k,i)$ e $(\nu P)_{i,j} = \sum_{k=1}^{r} \nu(k,i)p(i,j)$.

Prova Pelo Teorema 4.3, $I(\nu) = \sup_{u\gg 0} \sum_{l=1}^{s} \nu_l \ln\left(\frac{u_l}{(uQ)_l}\right)$, sendo $s = r^2$, $u = (u_1, \cdots, u_s)$, $\nu = (\nu_1, \cdots, \nu_s)$ e $Q = \{q(i,j)\}_{i,j=1}^{s}$. Observe que u tem todas entradas positivas, $\nu \in \mathcal{M}(\mathcal{T})$, com $T = \{1, \cdots, s\}$, e Q é a matriz de transição da Cadeia $\{Z_n\}_{n\geq 1}$. Outra maneira de escrever $I(\nu)$ é assumir u e ν como matrizes de ordem $r \times r$, isto é, $u = \{u(i,j)\}_{i,j=1}^{r}$ e $\nu = \{\nu(i,j)\}_{i,j=1}^{r}$. Agora $\nu \in \mathcal{M}(\mathcal{S}^{\in})$ e $u \in \{(x,y) \in \mathbf{R}^2 : x > 0,\, y > 0\}$. Portanto,

$$I(\nu) = \sup_{u\gg 0} \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{\nu(i,j)}{\sum_{k=1}^{r} u(k,i)p(i,j)}\right).$$

Observe que $(uQ)_l = \sum_{k=1}^{r} u(k,i)p(i,j)$ para um determinado par $(i,j) \in S^2$. Por exemplo, considerando o Exemplo 5.1,

$$\begin{aligned}
\frac{u_3}{(uQ)_3} &= \frac{u_3}{\sum_{m=1}^{4} u_m q(m,3)} = \frac{u_3}{u_1 0 + u_2 \frac{1}{3} + u_3 0 + u_4 \frac{1}{3}} \\
&= \frac{u(2,1)}{u(1,1)0 + u(1,2)p(2,1) + u(2,1)0 + u(2,2)p(2,1)} = \frac{u(2,1)}{\sum_{k=1}^{2} u(k,2)p(2,1)} .
\end{aligned}$$

Assuma que ν não é invariante por transição. Então, $\nu_c(j_0) < \nu_l(j_0)$ para algum $j_0 \in \{1, \cdots, r\}$. Para $i \in \{1, \cdots, r\}$ e $\alpha \in \mathbf{R}$ tome u tal que $u(i,j) = 1$, quando $j \neq j_0$, e $u(i,j_0) = e^{\alpha}$. Assim, $u(\cdot, j) = e^{\alpha}$, se $j = j_0$, e $u(\cdot, j) = 1$, se $j \neq j_0$. Prosseguindo,

$$\sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)}{\sum_{k=1}^{r} u(k,i)p(i,j)}\right) = \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{u(1,j)}{ru(1,i)p(i,j)}\right)$$

$$\sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{1}{rp(i,j)}\right) + \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{u(1,j)}{u(1,i)}\right) , \tag{5.2}$$

pois $u(k,i)$ é constante para todo k e também $u(i,j)$ é constante para todo i. Uma vez que

$$\frac{u(1,j)}{u(1,i)} = \begin{cases} \frac{1}{1}, & \text{se } j \neq j_0 \text{ e } i \neq j_0 \\ \frac{e^{\alpha}}{e^{\alpha}}, & \text{se } j = j_0 \text{ e } i = j_0 \\ \frac{1}{e^{\alpha}}, & \text{se } j \neq j_0 \text{ e } i = j_0 \\ \frac{e^{\alpha}}{1}, & \text{se } j = j_0 \text{ e } i \neq j_0 \end{cases} ,$$

(5.2) fica sendo

$$- \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(rp(i,j)\right) + \sum_{j=1}^{r} \nu(j_0, j)(-\alpha) + \sum_{i=1}^{r} \nu(i, j_0)\alpha =$$

$$- \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(rp(i,j)\right) + \alpha\left(\nu_l(j_0) - \nu_c(j_0)\right) , \tag{5.3}$$

Como $\sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(rp(i,j)\right)$ é limitado, então $\sup_{u \gg 0} \sum_{j=1}^{r} \sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{\mu(i,j)}{\sum_{k=1}^{r} \mu(k,i)p(i,j)}\right) = +\infty$, pois $\lim_{\alpha \to +\infty} \alpha\left(\nu_l(j_0) - \nu_c(j_0)\right) = +\infty$.

Agora vamos assumir que ν é invariante por transição. Para todo $u \gg 0$,

$$\sum_{i=1}^{r} \sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{\sum_{k=1}^{r} u(k,i)\nu_l(j)}{\sum_{k=1}^{r} u(k,j)\nu_c(i)}\right) = \sum_{i=1}^{r} \sum_{j=1}^{r} \nu(i,j) \ln\left(\sum_{k=1}^{r} u(k,i)\nu_l(j)\right)$$

$$\sum_{i=1}^{r} \sum_{j=1}^{r} \nu(i,j) \ln\left(\sum_{k=1}^{r} u(k,j)\nu_c(i)\right) = \sum_{i=1}^{r} \sum_{j=1}^{r} \nu(i,j) \ln\left(\sum_{k=1}^{r} u(k,i)\nu_l(j)\right)$$

$$\sum_{i=1}^{r} \sum_{j=1}^{r} \nu(i,j) \ln\left(\sum_{k=1}^{r} u(k,i)\nu_c(j)\right) = 0, \tag{5.4}$$

pois $\nu_c(j) = \nu_l(j)$.
De (5.4),

$$\sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)\nu_l(j)}{\sum_{k=1}^{r} u(k,j)\nu(i,j)}\right) = \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)\nu_c(i)}{\sum_{k=1}^{r} u(k,i)\nu(i,j)}\right) \quad (5.5)$$

A seguir defina as seguintes entropias relativas, de acordo com a Definição 2.16(ii):

$$\begin{aligned} H\left(\nu_c(\cdot|i)|p(i,\cdot)\right) &= \sum_{j=1}^{r} \nu_c(j|i) \ln\left(\frac{\nu_c(j|i)}{p(i,j)}\right), \ \nu_c(j|i) = \frac{\nu(i,j)}{\nu_c(i)}; \\ H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) &= \sum_{i=1}^{r} \nu_l(i|j) \ln\left(\frac{\nu_l(i|j)}{u(i|j)}\right), \ u(i|j) = \frac{u(i,j)}{\sum_{k=1}^{r} u(k,j)}. \end{aligned}$$

Segue que

$$I(\nu) - \sum_{i=1}^{r} \nu_c(i) H\left(\nu_c(\cdot|i)|p(i,\cdot)\right) = \sup_{u\gg 0} \sum_{j=1}^{r}\sum_{i=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)}{\sum_{k=1}^{r} u(k,i)p(i,j)}\right)$$

$$- \sum_{i=1}^{r} \nu_c(i) \sum_{j=1}^{r} \frac{\nu(i,j)}{\nu_c(i)} \ln\left(\frac{\nu(i,j)}{\nu_c(i)p(i,j)}\right) = \sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)}{\sum_{k=1}^{r} u(k,i)p(i,j)}\right)$$

$$- \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \frac{\nu(i,j)}{\nu_c(i)p(i,j)} = \sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)\nu_c(i)}{\sum_{k=1}^{r} u(k,i)\nu(i,j)}\right)$$

$$= \sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)\nu_l(j)}{\sum_{k=1}^{r} u(k,i)\nu(i,j)}\right), \quad (5.6)$$

sendo a última igualdade justificada por (5.5). Continuando a desenvolver (5.6),

$$\begin{aligned} &\sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i,j)\nu_l(j)}{\sum_{k=1}^{r} u(k,i)\nu(i,j)}\right) = \sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} \nu(i,j) \ln\left(\frac{u(i|j)}{\nu_l(i|j)}\right) \\ = &\sup_{u\gg 0} \sum_{i=1}^{r}\sum_{j=1}^{r} -\nu(i,j) \ln\left(\frac{\nu_l(i|j)}{u(i|j)}\right) = \sup_{u\gg 0} \left\{ -\sum_{j=1}^{r} \nu_l(j) H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) \right\}. \end{aligned}$$

Portanto, concluímos que

$$I(\nu) - \sum_{i=1}^{r} \nu_c(i) H\left(\nu_c(\cdot|i)|p(i,\cdot)\right) = \sup_{u\gg 0} \left\{ -\sum_{j=1}^{r} \nu_l(j) H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) \right\}. \quad (5.7)$$

Agora observe que (5.7) é sempre não positiva, pois $-\sum_{j=1}^{r} \nu_l(j) H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) \leq 0$, uma vez que $H\left(\nu_l(\cdot|j)|u(\cdot|j)\right)$ é função não negativa, conforme Teorema 2.5 (ii). Além disso,

$H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) = 0$ somente quando $u = \nu$. Logo,

$$\sup_{u \gg 0} \left\{ -\sum_{j=1}^{r} \nu_l(j) H\left(\nu_l(\cdot|j)|u(\cdot|j)\right) \right\} = 0\,,$$

e isto quer dizer que $I(\nu) = \sum_{i=1}^{r} \nu_c(i) H\left(\nu_c(\cdot|i)|p(i,\cdot)\right)$. $\square$

GIACOMELLI, M.A. Large Deviations Principle in Finite and Homogeneous Markov Chains. *Rev. Mat. Estat.* (São Paulo), v. 20, p. 125-153, 2002.

- *ABSTRACT: Cramer-Chernoff's Large Deviations Principle (LDP) was create in 1952 for sequence of random variables that are independent and identically distributed. Extensions it has been made for random variables with some kind of stochastic dependence. This paper is restricted to the context of finite and homogeneous Markov chains. Some definitions and results are given initially for use later. The objective is to analyse three situations where LDP is applied: the sample mean, the empirical measure and the pair empirical measure. The central reference is Ellis (1985), who calls mean, empirical measure and pair empirical measure by level 1, level 2 and level 3 sums, respectively.*

- *KEYWORDS: Large Deviations Principle, sample mean in Markov Chains, empirical measure in Markov Chains.*

Referências

ACKER, F.; DICKSTEIN, F. *Uma introdução à análise convexa.* In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 14., 1983 Poços de Caldas, 163p.

BUCKLEW, J. A. *Large deviations techniques in decision, simulation and estimation.* New York: Wiley, 1990. 320p.

CHERNOFF, H. A Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math Statist* . v.23, p.493-507, 1952.

CRAMÉR, H. Sur un nouveau théoreme-limit de la théorie des probabilités. In: COLLOQUE CONSACRÉ à LA THÉORIE

DES PROBABILITÉ, 1937, Paris. *Actualités scientifiques et industrielles*. v.336, p.5-23.

DEMBO, A.; ZEITOUNI, O. *Large deviations*: techniques and applications. London: Jones and Bartlett, 1993. 346p.

DONSKER, M. D.; VARAHDAN, S.R.S. Asymptotic evaluation of certain markov process expectations for large time I. *Comm. Appl Math.*, v.28, p.1-47, 1975.

DONSKER, M. D.;VARAHDAN, S.R.S. Asymptotic evaluation of certain markov process expectations for large time II. *Comm. Appl Math.*, v.28, p.279-301, 1975.

DONSKER, M. D.; VARAHDAN, S.R.S. Asymptotic evaluation of certain markov process expectations for large time III. *Comm. Appl Math.*, v.29, p.389-461, 1976.

DONSKER, M. D. VARAHDAN, S.R.S. Asymptotic evaluation of certain markov process expectations for large time IV. *Comm. Appl Math.*, v.36, p.183-212, 1983.

ELLIS, R. S. *Entropy, large deviations and statistical mechanics.* New York: Springer-Verlag, 1985. 364p.

FREIDLIN, M. I. ; WENTZELL, A. D. *Random perturbations of dynamival systems* New York: Springer-Verlag, 1984. 413p.

GÄRTNER, J. On Large deviations from the invariant measure. *Prob. Appl.*, v.22, p.24-39, 1977.

TAYLOR, H. *A First course in stochastic processes* . 2nd ed. New York: Academic Press, 1975. 557p.

KEMENY, J. G. ; SNELL, J. L. *Finite markov chains* . New Jersey: Van Nostrand-Princeton, 1960. 210p.

LANCASTER, P. *Theory of matrices.* New York: Academic Press, 1969. 570p.

LANFORD, O. C. Entropy and equilibrium states in classical statistical mechanics. In: Lenard, A. (Ed.). *Statistical mechanics and mathematical problems.* New York: Springer-Verlag, 1973. p.1-113. (Lecture notes in phisics, v.20).

LIGGET, T. M. *Interacting particle systems* . New York: Springer-Verlag, 1985. 490p.

LIMA, E. L. *Curso de Análise* . Rio de Janeiro: Instituto de Matemática Pura e Aplicada (IMPA)-CNPq, 1982. v. 1, 344p.

SANOV, I. N. On the probability of large deviations of random variables. In: SELECTED TRANSLATIONS IN MATHEMATICAL STATISTICS AND PROBABILITY, I. New York: 1961, p.213-244.

SENETA, E. *Non-negative matrices and markov chains.* New York: Springer-Verlag, 1981. 279p.

VARADHAN, S. R. S. *Large deviations and applications.* Philadelphia: SIAM, 1984. 340p.

VARES, M. E. *Grandes desvios em processos markovianos.* In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 15., 1985, Rio de Janeiro. 136p.

Recebido em 18.09.2000.