

**GRANDES DESVIOS  
EM PROCESSOS MARKOVIANOS**

**Maria Eulália Vares**

COPYRIGHT © - 1985 - by Maria Eulália Vares

Nenhuma parte deste livro pode ser reproduzida,  
por qualquer processo, sem a permissão da autora.

INSTITUTO DE MATEMÁTICA PURA E APLICADA

Estrada Dona Castorina, 110

22.460 - Rio de Janeiro - RJ

A

Edson e Juliana



Grandes Desvios em Processos Markovianos - 15º Colóquio Brasileiro de Matemática

Maria Eulália Vares

E R R A T A

(a) Duas correções mais importantes:

(i) No Lema 1.4 (p. 22 e 23) devemos tomar  $C$  aberto e convexo, e a obs. (p. 23 linha 4) deve ser: "Se  $f(n) > -\infty$  para algum  $n$ , então  $\exists n^* < +\infty$  t.q.  $f(m) > -\infty \forall m \geq n^*$ . (Este fato é utilizado para concluirmos que  $f(m)/m$  converge.)"

(ii) No final do "esboço da prova" do Teorema 1.13 há um erro: obviamente a aproximação utilizada (p. 44, linhas 1,2 e 3) não serve. Devemos ter cuidado para aproximarmos  $\log g$  por funções borelianas e limitadas de modo conveniente. (Ref. Donsker-Varadhan (1975).)

(b) Outras correções menos importantes:

| <u>Pág.</u> | <u>linha</u> | <u>onde se lê</u>                     | <u>leia-se</u>                        |
|-------------|--------------|---------------------------------------|---------------------------------------|
| 3           | 2            | $I(A)$                                | $-I(A)$                               |
| 15          | 3            | $np_i e^{sx_i} / \sum_j p_j e^{sx_j}$ | $np_i e^{sx_i} / \sum_j p_j e^{sx_j}$ |

| Pág.    | linha              | onde se lê                                                                                         | leia-se                                                                       |
|---------|--------------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| 16      | 6                  | $av+b \geq 1$                                                                                      | $av+b \geq 1$ e $b \in I$                                                     |
|         | -3                 | $u(t)$                                                                                             | $\hat{u}(t)$                                                                  |
| 17      | -1                 | $u(t)$                                                                                             | $\hat{u}(t)$                                                                  |
| 18      | -3                 | §1.1                                                                                               | Cap. 1                                                                        |
| 20      | -5                 | $\inf_{v \in A} I_{\pi}(v)$                                                                        | $-\inf_{v \in A} I_{\pi}(v)$                                                  |
| 23      | 1                  | $X_{n+m} =$                                                                                        | $\bar{X}_{n+m} =$                                                             |
|         | -4                 | $\lambda_p(y)$                                                                                     | $\lambda_{\mu}(y)$ **                                                         |
| 30      | 9                  | item (3)                                                                                           | item (4)                                                                      |
|         | 10                 | $t[x, +\infty) \leq t[x] \vee t(x, +\infty)$                                                       | $\bar{t}[x, +\infty) \leq \bar{t}[x] \vee t(x, +\infty)$                      |
| 31      | -2                 | (3)                                                                                                | (4)                                                                           |
| 36      | 8 e 9              | [eliminar final da frase, após a palavra "idêntica", pois C é <u>aberto e convexo</u> no Lema 1.4] |                                                                               |
| 38      | 1                  | inf                                                                                                | sup                                                                           |
| 39      | 6                  | 1975                                                                                               | 1961                                                                          |
| 41      | 11                 | $-I_{\pi}(A)$                                                                                      | $-I_{\pi}(A)$                                                                 |
| 47      | 15                 | $\int(x) u(dx)$                                                                                    | $\int x  u(dx)$                                                               |
| 48      | 5                  | $\Lambda(\bar{A})$                                                                                 | $-\Lambda(\bar{A})$                                                           |
| 54      | 9                  | $\epsilon^{-1} f(x_0)$                                                                             | $-\epsilon^{-1} f(x_0)$                                                       |
| 57      | 5                  | $s > 0$                                                                                            | $0 < s < +\infty$                                                             |
|         | 7                  | [acrescentar "se $\epsilon \leq \epsilon_0$ "]                                                     |                                                                               |
| 58      | 5                  | $p > 0$                                                                                            | $0 < s < +\infty$                                                             |
| 60 e 61 | sempre que ocorrer | $\int_0^T \langle \varphi(s), dW_s \rangle$                                                        | $\int_0^T \langle \hat{\varphi}(s), dW_s \rangle$                             |
| 62      | 1                  | $\geq \delta$                                                                                      | $> \delta$                                                                    |
|         | -2                 | $\frac{8d^2}{\delta} \sqrt{\frac{NT}{\pi}} \epsilon e^{-s_0/2\epsilon^2} \leq$                     | $\frac{8d^2}{\delta} \sqrt{\frac{NT}{\pi}} \epsilon e^{-s_0/\epsilon^2} \leq$ |
|         |                    | $\leq \frac{1}{2} e^{-\frac{1}{2} \frac{s-y}{\epsilon^2}}$                                         | $\leq \frac{1}{2} e^{-\frac{s-y}{\epsilon^2}}$                                |
|         | -1                 | 4                                                                                                  | 8                                                                             |
| 63      | 2                  | $\frac{\epsilon}{2} \sum_{k=1}^N N  W_{kT/N} - W_{(k-1)T/N} ^2$                                    | $\frac{\epsilon^2}{2} \sum_{k=1}^N \frac{N}{T}  W_{kT/N} - W_{(k-1)T/N} ^2$   |
|         | 3                  | $\sqrt{N} (W_{kT/N} - W_{(k-1)T/N})$                                                               | $\sqrt{\frac{N}{T}} (W_{kT/N} - W_{(k-1)T/N})$                                |
|         | 7                  | $\frac{\epsilon}{2} \sum_{i=1}^{dN}  Z_i ^2$                                                       | $\frac{\epsilon^2}{2} \sum_{i=1}^{dN}  Z_i ^2$                                |
|         | -2                 | $c_{\theta}^{dN} e^{-s/\epsilon^2} e^{y/\epsilon^2}$                                               | $C_{\theta}^{dN} e^{-s/\epsilon^2} e^{y/2\epsilon^2}$                         |

| Pág. | linha   | onde se lê                                     | leia-se                                                                                                  |
|------|---------|------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| 63   | -2      | $c_{\theta}^{dN} e^{\gamma/\epsilon^2}$        | $C_{\theta}^{dN} e^{\gamma/2\epsilon^2}$                                                                 |
| 70   | -1      | D                                              | $\bar{D}$                                                                                                |
| 73   | 9       | $B_{2\mu}(0)$                                  | $\overline{B_{2\mu}(0)}$                                                                                 |
|      | -1      | $\tau_1$<br>[a acrescentar "etc..."]           | $\eta_1$                                                                                                 |
| 78   | 3       | $\tau_{\alpha}$                                | $R_{\alpha}$                                                                                             |
| 79   | 9       | d/10L                                          | d/20L                                                                                                    |
|      | 11      | $\leq \mu$                                     | $< \mu/2$                                                                                                |
|      | 13      | condição                                       | condição (I)                                                                                             |
|      | 15      | $y_1 \notin \bar{D}$ , com $ y_1 - y_0  = \mu$ | $y_1$ t.q. $ y_1 - y_0  = \min_{y \in \bar{D}}  y_1 - y  = \mu$ ,<br>onde diminuimos $\mu$ se necessário |
| 80   | 9       | $T_1 - t_1 + 3\mu \leq$                        | $T_1 - t_1 + 3\mu \leq t \leq$                                                                           |
|      | 10      | y                                              | $y_1$                                                                                                    |
|      | 12 e 16 | $< \mu$                                        | $< \mu/2$                                                                                                |
| 82   | 11      | T                                              | $t_0$                                                                                                    |
| 107  | 14      | $y \in \partial G$                             | $y \in \partial G$                                                                                       |
| 108  | 13      | Lipschitz                                      | continuidade                                                                                             |
| 109  | 2       | $\delta$                                       | $\eta$                                                                                                   |
| 110  | -4      | (3.6)                                          | (3.6a)                                                                                                   |
| 113  | 14      | $(t+s)\gamma$                                  | $(t+s)\gamma_{\epsilon}$                                                                                 |
| 117  | 7       | $b_0(x)$                                       | $b(x)$                                                                                                   |
| 119  | 12      | $B_c'(q)$                                      | $B_c(q)$                                                                                                 |
| 121  | 15/16   | Pianigiari                                     | Pianigiani                                                                                               |
| 125  | 8       | §3.5                                           | §4.5                                                                                                     |
| 126  | 2       | $\alpha$                                       | $\mu$                                                                                                    |
| 129  | 8       | $\mu_0$                                        | $\mu_g$                                                                                                  |
|      | 8       | $\tau_{\epsilon}(X)/\epsilon$                  | $\epsilon \tau_{\epsilon}(X)$                                                                            |
| 130  | 5       | $\in$                                          | $\notin$                                                                                                 |





Í N D I C E

|                                                                                                               | <u>Pág.</u> |
|---------------------------------------------------------------------------------------------------------------|-------------|
| INTRODUÇÃO .....                                                                                              | 1           |
| CAPÍTULO I                                                                                                    |             |
| Teoremas de Grandes Desvios à Cramèr-Chernoff.....                                                            | 9           |
| §1.1 Grandes desvios com relação à lei dos grandes números em $\mathbb{R}$ : Teorema de Cramèr-Chernoff ..... | 9           |
| §1.2 Extensões do Teorema de Cramèr-Chernoff .....                                                            | 19          |
| §1.3 Teorema de Sanov. Grandes desvios de distribuições empíricas .....                                       | 39          |
| Comentários e Referências adicionais .....                                                                    | 44          |
| Exercícios .....                                                                                              | 46          |
| CAPÍTULO 2                                                                                                    |             |
| Pequenas Perturbações Aleatórias de Sistemas Dinâmicos - Teoria de Freidlin e Wentzell - .....                | 49          |
| §2.1 Introdução do modelo. Estimativas básicas.....                                                           | 49          |
| §2.2 Saída de um domínio.....                                                                                 | 68          |
| EXERCÍCIOS .....                                                                                              | 91          |

|                                                                                                                        | <u>Pág.</u> |
|------------------------------------------------------------------------------------------------------------------------|-------------|
| CAPÍTULO 3                                                                                                             |             |
| Metaestabilidade e Grandes Desvios .....                                                                               | 96          |
| §3.1 Comentários Gerais .....                                                                                          | 96          |
| §3.2 Comportamento, metaestável de uma classe de sistemas dinâmicos sujeitos a pequenas perturbações estocásticas..... | 100         |
| §3.3 Comportamento metaestável em sistemas com dinâmica determinística.....                                            | 121         |
| REFERÊNCIAS.....                                                                                                       | 132         |

## §0. Introdução

A obtenção de aproximações para probabilidades de "grandes desvios", ou seja, de certos "eventos raros" (probabilidade tendendo a zero) tem desempenhado um importante papel em diferentes aspectos da teoria de probabilidades e suas aplicações.

No contexto clássico os primeiros resultados neste sentido foram obtidos por Cramèr, tratando de grandes desvios com relação à lei dos grandes números. Posteriormente seu trabalho foi ampliado por Chernoff (1952), e ainda com contribuições de muitos outros autores. Este tipo de resultados constituiu-se então, conjuntamente com teoremas do tipo "Lei dos Grandes Números" e "Teorema Central do Limite" no corpo básico de resultados para o estudo de problemas assintóticos no contexto de seqüências de variáveis aleatórias independentes e com mesma lei, permitindo um desenvolvimento de métodos assintóticos em Estatística.

Por outro lado uma razão igualmente forte para o estudo de "grandes desvios" em situações de dependência entre as variáveis aleatórias também já estava latente há algum tempo. Ela aparece naturalmente na Mecânica Estatística, na tentativa de justificar a descrição das propriedades macroscópicas de sistemas termodinâmicos em equilíbrio, através de um número reduzido de variáveis (e.g., temperatura, pressão, etc.), apesar da enormidade do sistema. Tais sistemas constituem-se de um número muito grande ( $\approx 10^{23}$ ) de componentes ("moléculas", "átomos") que interagem de modo complicado. Lanford (1961), voltado principalm

te para este problema da Mecânica Estatística, apresentou idéias básicas para a extensão (em várias direções) dos resultados de Cramèr - Chernoff. Estas idéias, desenvolvidas posteriormente por Bahadur e Zabell (1979), também favorecem um melhor entendimento do papel da "entropia" na teoria de grandes desvios e as conseqüentes aplicações.

Voltando ao contexto anteriormente descrito, esboçemos brevemente os resultados de Cramèr - Chernoff. Para isto, sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes e de mesma lei em certo espaço de probabilidade  $(\Omega, \mathcal{G}, P)$ . Se  $X_1$  for integrável e  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  ( $n \geq 1$ ), a lei fraca dos grandes números diz que  $P(\bar{X}_n \in A) \rightarrow 0$  quando  $A$  for um boreliano de  $\mathbb{R}$  tal que  $m \stackrel{\text{def}}{=} E X_1 \notin \bar{A}$ , e o evento  $[\bar{X}_n \in A]$  representa uma situação de grande desvio com relação à lei dos grandes números. O problema consiste em obter estimativas mais precisas de  $P(\bar{X}_n \in A)$ , quando  $n \rightarrow +\infty$ . A situação típica no estudo de Cramèr - Chernoff é aquela em que tais probabilidades decaem exponencialmente a zero i.e., para uma ampla classe de  $A$ 's tais que  $E X_1 \notin \bar{A}$  tem-se

$$P(\bar{X}_n \in A) \approx e^{-nI(A)} \quad (0.1)$$

sendo que  $I(A) \in (0, +\infty]$ . O fato de termo  $I(A) > 0$  está ligado à hipótese de  $E e^{sX_1} < +\infty$  para  $s$  em alguma vizinhança do zero. Pode dizer que sob as condições mais gerais de Chernoff, o símbolo " $\approx$ " em (0.1) deve ser interpretado como equivalência logarítmica apenas, i.e.,

$$\frac{1}{n} \log P(\bar{X}_n \in A) \rightarrow -I(A) \quad (0.2)$$

onde  $I(A)$  é o que se chama de "entropia" em  $A$ . De fato, Chernoff (1952) mostra que supondo-se  $X_1$  integrável então (0.2) é válido para todo  $A$  do tipo  $(-\infty, x]$  ou  $[x, +\infty)$  sendo  $I(A) \in [0, +\infty]$  explicitamente caracterizado por

$$I(A) = \inf_{x \in A} \lambda(x),$$

onde

$$\lambda(x) = \sup_{t \in \mathbb{R}} (tx - \log E e^{tX_1}).$$

O nome "grande desvio" dado a eventos do tipo  $[\bar{X}_n \in A]$  onde  $m \notin \bar{A}$ , fica de certa forma justificado pelo contraste com as situações descritas pelo Teorema Central do Limite: no caso de  $\sigma^2 = \text{Var } X_1 < +\infty$  o T.C.L. diz

$$P(\bar{X}_n \in m + \frac{1}{\sqrt{n}} B) \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \int_B e^{-u^2/2\sigma^2} du$$

para todo boreliano  $B$ . Se  $0 \notin \bar{B}$  e portanto  $A_n \stackrel{\text{def}}{=} m + \frac{1}{\sqrt{n}} B$  está a uma distância positiva ( $\geq c/\sqrt{n}$ ) de  $m$ , os eventos  $[\bar{X}_n \in A_n]$  representam um "pequeno" ou "moderado desvio" em relação à lei dos grandes números, e suas probabilidades tendem a um limite não nulo, se  $B$  tiver medida de Lebesgue positiva.

O problema considerado por Cramer e Chernoff com relação às variáveis  $\bar{X}_n$  acima definidas tem uma generalização natu-

ral para famílias de processos estocásticos. Exemplificando tal situação, nos restringiremos a uma classe bem particular de processos, obtidos via uma perturbação aleatória de certos sistemas dinâmicos determinísticos. Trata-se aqui (Wentzell e Freidlin (1970)) de uma família de processos de difusão  $X_\epsilon(\cdot)$ ,  $\epsilon > 0$ , definidos por equação diferencial estocástica

$$\begin{aligned} dX_\epsilon(t) &= b(X_\epsilon(t))dt + \epsilon dW_t \\ X_\epsilon(0) &= x \end{aligned} \tag{0.3}$$

onde  $(W_t)$  é um Movimento Browniano; isto representa então uma "pequena perturbação" aleatória do sistema dinâmico

$$\begin{aligned} \dot{X}(t) &= b(X(t)) \\ X(0) &= x, \end{aligned} \tag{0.4}$$

onde estamos supondo  $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$  de modo a garantir existência e unicidade de solução (nos sentidos usuais) em (0.3) e (0.4), e estamos interessados em  $\epsilon$  tendendo a zero.

Fixado um intervalo limitado  $[0, T]$ , se observássemos  $X_\epsilon(t)$  e  $X(t)$  para  $0 \leq t \leq T$ , com grande probabilidade quando  $\epsilon$  é pequeno, veríamos  $X_\epsilon(\cdot)$  oscilando rapidamente em torno a  $X(\cdot)$ ; com efeito, é fácil constatar que dado  $\delta > 0$

$$P(\sup_{t \leq T} |X_\epsilon(t) - X(t)| > \delta) \xrightarrow{\epsilon \downarrow 0} 0 \text{ (exponencialmente)}. \tag{0.5}$$

Ora, isto é exatamente análogo à situação considerada por Cramér-Chernoff, substituindo-se  $\bar{X}_n$  e  $m$  por  $X_\epsilon(\cdot)$  e  $X(\cdot)$ . De fato, pensando em  $X_\epsilon(\cdot)$  como variáveis aleatórias tomando valores no espaço  $C[0, T] \stackrel{\text{def}}{=} \{f = [0, T] \rightarrow \mathbb{R}^d \mid f \text{ é contínua}\}$

munido da métrica  $\rho_T(\eta, \xi) = \sup_{t \in [0, T]} |\eta(t) - \xi(t)|$  então (0.5) pode ser reescrita como

$$P(\rho_T(X_\epsilon, X) > \delta) \longrightarrow 0 \quad (\text{exponencialmente}), \quad (0.6)$$

para cada  $\delta > 0$ , quando  $\epsilon \downarrow 0$ . (Isto também justifica o nome "pequenas perturbações".)

Coloca-se então a questão referente ao comportamento de  $P(X_\epsilon \in A)$  quando  $\epsilon \downarrow 0$ , onde  $A$  é um boreliano de  $C[0, T]$ , a uma distância positiva da trajetória determinística  $(X(t): 0 \leq t \leq T)$  sendo  $0 < T < +\infty$  fixado (porém arbitrariamente).

À primeira vista esta preocupação em estimar a probabilidade de eventos raros pode parecer estranha. Entretanto tais estimativas serão fundamentais na determinação do comportamento típico de  $X_\epsilon(\cdot)$  durante intervalos de tempo que podem se tornar arbitrariamente longos (quando  $\epsilon \downarrow 0$ ). Neste caso (0.5) deixa de ser útil; possivelmente  $X_\epsilon(\cdot)$  terá um comportamento drasticamente diferente de  $X(\cdot)$ , i.e., grandes desvios poderão, e deverão ocorrer. Para determinarmos o comportamento típico será necessário sabermos quais entre tais eventos raros são mais, ou menos, "improváveis". Exemplificando, consideremos o caso em que existam um domínio limitado  $D$ , invariante para (0.4), e um ponto  $x_0 \in D$  para o qual convergem todas as trajetórias de (0.4) que partem de algum lugar em  $\bar{D}$ . Já para  $\epsilon > 0$ , as trajetórias de  $X_\epsilon(\cdot)$  escaparão de  $D$  após um tempo aleatório  $\tau_\epsilon < +\infty$ , com probabilidade 1. Algumas perguntas naturais e interessantes são: Qual é o tempo médio para a saída de  $D$ ? Como ocorre tal sai-

da? Por onde?

Estas são algumas das questões respondidas pela Teoria de Freidlin e Wentzell(1970), envolvendo estimativas das probabilidades de grandes desvios para  $(X_\epsilon(t): 0 \leq t \leq T)$ , para cada  $T < \infty$ . Serão objeto de nosso estudo no Capítulo 2. Estas mesmas estimativas nos permitirão provar ainda um resultado interessante e bastante intuitivo sobre  $\tau_\epsilon$ . Devido à forte atração em direção a  $x_0$  que atua durante cada intervalo  $[S, T]$  enquanto  $X_\epsilon(\cdot)$  permanece em  $D$ , e como  $\tau_\epsilon$  deve tender a  $+\infty$  quando  $\epsilon \downarrow 0$ , é razoável conjecturar a perda de memória das variáveis  $\tau_\epsilon$  ( $\epsilon \downarrow 0$ ). Isto será formalizado no Capítulo 3 mostrando-se que, existem constantes  $\beta_\epsilon / +\infty$  tais que as variáveis aleatórias  $\tau_\epsilon / \beta_\epsilon$  convergem em lei a uma variável aleatória com distribuição exponencial, quando  $\epsilon \downarrow 0$ .

Na verdade, a redação destas notas sobre a Teoria de Freidlin e Wentzell bem como problemas de grandes desvios de modo mais geral, foi fundamentalmente motivada pelo que discutiremos no Capítulo 3. Mais uma vez, esta motivação vem de um problema ligado à Mecânica Estatística. Mecânica Estatística que pode ser considerada como uma das raízes da própria teoria de processos estocásticos, através do trabalho de Boltzmann e das discussões sobre o aspecto probabilístico de seu célebre Teorema H.

Na situação em estudo refiro-me de modo específico ao problema do entendimento e modelagem do fenômeno da metaestabilidade. Tipicamente poderíamos associar comportamento metaestável à seguinte descrição informal: um processo estocástico que



possue uma única medida de probabilidade estacionária, mas que se comporta como se estivesse em equilíbrio descrito por outra medida invariante. Após um "tempo muito longo" ocorre a transição para o verdadeiro equilíbrio; uma característica comum é o modo abrupto e imprevisível como tal transição ocorre. O fenômeno físico da metaestabilidade ocorre, na natureza, em muitas situações de transição de fase. Para uma discussão sobre os aspectos mais físicos ligados à fenomenologia da metaestabilidade referimos a [29]. M. Cassandro, A. Galves, E. Olivieri e M. Vares (1984) propuseram modelar tal situação através do comportamento típico das médias temporais das trajetórias, pois são estas as estatísticas mais naturais de um processo estocástico. As características marcantes do que chamamos de "comportamento metaestável" seriam então a aparente estabilidade de médias temporais tomadas ao longo das trajetórias, seguindo-se uma transição brusca e imprevisível, e finalmente a estabilidade (ou ainda repetições destas situações de aparente estabilidade até estabilizar-se, finalmente). Do ponto de vista probabilístico isto pode ser facilmente interpretado quando a transição é fruto da ocorrência de uma "grande flutuação" requerendo uma sucessão de eventos raros. Isto determinará a forma abrupta e imprevisível como ocorre: é a falta de memória, associada à distribuição exponencial. Por outro lado há uma outra escala de tempo, bem mais curta, que permite a "terminalização" do sistema i.e., permite que as médias temporais tomadas nesta escala se estabilizem, dando a impressão de que o sistema está em equilíbrio. Sendo pois um fenômeno de "grandes flutuações" pode-se antever a utilização de teorias de grandes desvios

na verificação deste tipo de comportamento em modelos estocásticos. Um pouco disto será discutido no Capítulo 3.

Quanto à disposição dos assuntos neste curso, cabem algumas observações. No Capítulo 1 tratamos de teoremas de grandes desvios para somas de variáveis aleatórias independentes e de mesma lei. O caso clássico de Cramèr e Chernoff refere-se a variáveis reais. Generalizações a espaços infinito dimensionais (Bahadur-Zabell (1979), Donsker-Varadhan (1975)) fornecem aplicações importantes para a Estatística (como é o caso do Teorema de Sanov que discutiremos) mas também para a Teoria de processos estocásticos, permitindo unificar estes resultados com a Teoria de Freidlin e Wentzell, discutida no Capítulo 2. Entretanto, na forma como as notas foram escritas, os Capítulos 2 e 3 são independentes do primeiro. No final do Capítulo 3 faremos uma breve discussão sobre modelos com dinâmica determinística e que apresentam um comportamento "metaestável" (cf. [22,28,34].)

Como comentário final, devo ressaltar que o presente texto não pretende fazer um resumo completo do que tem sido feito em "grandes desvios"; apenas discutir certas idéias básicas.

Meu agradecimento a Marzio Cassandro, Antônio Galves, Enzo Olivieri, Errico Presutti, Hermann Rost e Roberto Schonmann pelas longas e proveitosas discussões que tanto nos ensinam. Agradeço ainda a Lais Ventura Santos pela paciência e excelente trabalho de datilografia.

## CAPÍTULO 1

## TEOREMAS DE GRANDES DESVIOS À CRAMÈR - CHERNOFF

§1.1 Grandes desvios com relação à lei dos grandes números em  $\mathbb{R}$ :Teorema de Cramèr - Chernoff.

Sejam  $X_1, X_2, \dots$  variáveis aleatórias (reais) independentes, identicamente distribuídas e integráveis, e sejam  $m = E X_1 \in \mathbb{R}$ ,  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  para cada  $n \geq 1$ . A lei (forte) dos grandes números diz que  $\bar{X}_n \rightarrow m$  q.c. quando  $n \rightarrow +\infty$ . Em particular (lei fraca), se  $A$  for um boreliano de  $\mathbb{R}$  tal que  $m \notin \bar{A}$ , ( $\bar{A}$  = fecho de  $A$ ), então  $P(\bar{X}_n \in A) \rightarrow 0$  quando  $n \rightarrow +\infty$ .

Cramèr (1937) colocou-se então o problema de obter boas aproximações para tais probabilidades de "eventos raros" do tipo  $[\bar{X}_n \in A]$  quando  $A$  está a distância positiva de  $m$ .

A motivação inicial de Cramèr e Chernoff aparece ligada a problemas de Estatística Matemática, tais como determinar "eficiência" assintótica de um teste de hipóteses. Um exemplo bem simples, porém ilustrativo do que os motivou pode ser o seguinte: tem-se duas hipóteses  $H_0$  e  $H_1$ ; de modo que sob  $H_1$ ,  $X$  é variável aleatória de Bernoulli com probabilidade de sucesso  $p_1$  ( $i = 0, 1$ ), onde  $p_0 \neq p_1$  (s.p.g.  $p_1 > p_0$ ), i.e.,

$$P(X = 1 | H_i) = p_i = 1 - P(X = 0 | H_i) \quad i = 0, 1$$

onde  $0 < p_0 < p_1 < 1$ . Com base em  $n$  observações independen-

tes  $X_1, \dots, X_n$  da v.a.  $X$ , o teste da razão de verossimilhança tem a forma:

$$\text{rejeita } H_0 \text{ se } S_n \stackrel{\text{def}}{=} \sum_{i=1}^n X_i > k$$

para algum  $k$ . Para fins de comparação com outros testes, ou para se determinar quão grande deve ser a amostra, etc... faz-se necessário estimar as probabilidades de erro:

$$e_0 = P(S_n > k | H_0) \quad (\text{rejeitar } H_0 \text{ quando } H_0 \text{ é verdadeira})$$

$$e_1 = P(S_n \leq k | H_1) \quad (\text{não rejeitar } H_0 \text{ quando } H_0 \text{ é falsa}).$$

Cramèr e Chernoff estavam então interessados em boas aproximações para  $e_0, e_1$ . Uma prática comum talvez seja a utilização da aproximação normal. Isto é válido em situações tais como:  $n = 400, p_0 = 0,4, p_1 = 0,5, k = 180$ , etc. . Entretanto, quando  $n$  é muito grande de modo que a diferença das médias de  $S_n$  sob  $H_1$  e  $H_0$  ( $n(p_1 - p_0)$ ) é enorme quando comparada com o desvio padrão ( $\sqrt{np_0(1-p_0)}$ , sob  $H_0$  e  $\sqrt{np_1(1-p_1)}$  sob  $H_1$ ) precisamos olhar para caudas da distribuição de  $S_n$ , e aí a aproximação normal não funciona. Esta foi uma das razões que levaram Cramèr, e depois Chernoff, a estudarem o comportamento assintótico ( $n \rightarrow +\infty$ ) das caudas da função de distribuição de  $S_n = \sum_{i=1}^n X_i$  onde  $X_1, X_2, \dots$  são variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.). O trabalho inicial foi estendido em muitas direções. No curso discutiremos parte destas extensões, e no final da secção 1.1 o leitor encontra referên-

cias.

Os resultados iniciais (Cramèr) na verdade não se aplicam ao exemplo anterior pois tratavam de variáveis absolutamente contínuas. As hipóteses de Cramèr eram: (a)  $E e^{tX_1} < +\infty$  para todo  $t$  em algum intervalo  $(-\delta, \delta)$ ; (b)  $X_1$  com lei absolutamente contínua. Mais tarde (a) foi bastante relaxada (Esseen). Entretanto, os resultados de Cramèr são "exatos" i.e., obteve função  $f(n, a)$  tal que  $P(S_n \leq na)/f(n, a) \rightarrow 1$  quando  $n \rightarrow +\infty$  se  $a < E X_1$  (i.e.  $P(S_n \leq na) \sim f(n, a)$ , quando  $n \rightarrow +\infty$ ). Por outro lado, todos os resultados que discutiremos, referentes a problemas de grandes desvios darão apenas uma aproximação mais grosseira: equivalência logarítmica, ao invés de equivalência. No caso acima por exemplo, isto se expressa pelo cálculo de

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \leq na) = -\lambda(a).$$

Ou seja, estes resultados dizem apenas que  $\log P(S_n \leq na) \sim -n\lambda(a)$ . Denotaremos isto por  $P(S_n \leq na) \approx e^{-n\lambda(a)}$ , e chamaremos de equivalência logarítmica.

Começemos pois com o resultado de Cramèr (1937) e Chernoff (1952), segundo a formulação de Chernoff: nos diz que para  $X_1, X_2, \dots$  v.a.'s i.i.d. com  $E X_1 \in \mathbb{R}$ , então

$$P(\bar{X}_n \leq a) \approx m(a)^n \text{ se } a < E X_1$$

onde  $m(a)$  é o mínimo da função geradora de momentos de  $X_1 - a$ , i.e.,  $m(a) = \inf_t e^{-ta} E e^{tX_1}$ .

Antes de provarmos este resultado inicial em grandes desvios (bem clássico), modificaremos um pouco a notação.

Definição 1.1. Seja  $\mu$  uma probabilidade em  $\mathbb{R}$  e  $\hat{\mu}: \mathbb{R} \rightarrow (0, +\infty]$  sua transformada de Laplace (ou função geradora de momentos), i.e.,

$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{tx} \mu(dx). \quad (1.1)$$

A transformada de Cramèr de  $\mu$  é a função  $\lambda_{\mu}: \mathbb{R} \rightarrow [0, +\infty]$  definida por:

$$\lambda_{\mu}(a) = \sup_{t \in \mathbb{R}} (ta - \log \hat{\mu}(t)), \quad (1.2)$$

para todo  $a \in \mathbb{R}$ .

Observações:

1. Se  $\mu$  é a lei de  $X_1$ , então, na notação anterior,  $-\lambda_{\mu}(a) = \log m(a)$ .
2.  $\lambda_{\mu}(\cdot)$  é função convexa e semicontínua inferiormente (Cf. Exercício 1.1).

Teorema 1.2 (Cramèr (1937), Chernoff (1952)). Sejam  $X_1, X_2, \dots$  variáveis aleatórias i.i.d. . Seja  $\mu$  a distribuição comum e

seja  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $n \geq 1$ . Então:

- (a) Se  $E|X_1| < +\infty$  tem-se para todo  $n \geq 1$ :

$$P(\bar{X}_n \leq a) \leq e^{-n\lambda_\mu(a)}, \quad \text{quando } a \leq E X_1 \quad (1.3)$$

$$P(\bar{X}_n \geq a) \leq e^{-n\lambda_\mu(a)}, \quad \text{quando } a \geq E X_1 \quad (1.4)$$

(b) Para todo  $a \in \mathbb{R}$

$$-\lambda_\mu(a) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(\bar{X}_n \leq a) \quad (1.5)$$

$$-\lambda_\mu(a) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(\bar{X}_n \geq a). \quad (1.6)$$

Prova: (Cf. Chernoff (1952) ou Azencott (1980).)

(a) Basta provar que se  $E X_1 \geq 0$  então

$$P(\bar{X}_n \leq 0) \leq e^{-n\lambda_\mu(0)}, \quad (1.7)$$

pois os outros casos se reduzem facilmente a este. (Exercício 1.2). Mas

$$e^{-\lambda_\mu(0)} = \inf_{t \in \mathbb{R}} \hat{\mu}(t) = \inf_{t \leq 0} \hat{\mu}(t)$$

pois, pela desigualdade de Jensen  $\hat{\mu}(t) = E e^{tX_1} \geq e^{tEX_1} \geq 1 = \hat{\mu}(0)$  para  $t \geq 0$ . Ademais, se  $t \leq 0$  temos:

$$P(\bar{X}_n \leq 0) = P(S_n \leq 0) \leq P(e^{tS_n} \geq 1) \leq E e^{tS_n} = (\hat{\mu}(t))^n$$

provando (1.7).

(b) Novamente, basta provar (1.5) para  $a = 0$ . A prova de

Chernoff baseia-se no Lema 1.3 abaixo que é essencialmente o caso de v.a.'s discretas. Notemos antes que nos casos  $P(X_1 < 0) = 0$  ou  $P(X_1 > 0) = 0$ , (1.5) fica trivial. Por isto já supomos  $0 < P(X_1 < 0)$  e  $0 < P(X_1 > 0)$  no que segue.

Lema 1.3. Suponhamos  $r \geq 1$  e  $x_1, \dots, x_r$  reais tais que

$$\min_i x_i < 0 < \max_i x_i$$

e sejam  $p_i > 0$   $i = 1, \dots, r$ . Seja

$$\alpha = \inf_t \sum_i p_i e^{tx_i}$$

Então, existem  $c > 0$  e  $N \in \mathbb{N}$  tais que para todo  $n \geq N$  podemos encontrar  $n_i$  ( $i = 1, \dots, r$ ) inteiros positivos t.q.

$$\sum_{i=1}^r n_i = n$$

$$\sum_{i=1}^r n_i x_i \leq 0 \tag{1.8}$$

$$p(n_1, \dots, n_r) \stackrel{\text{def}}{=} n! \prod_{i=1}^r \frac{p_i^{n_i}}{n_i!} \geq c n^{-r/2} \alpha^n$$

(É claro que (1.8) implica a validade de (1.5) com  $a = 0$  no caso de  $\mu$  concentrada em conjunto finito.)

Prova: Pelá hipótese podemos tomar  $s \in \mathbb{R}$  tal que  $\alpha = \sum_i p_i e^{sx_i}$ .

Por Stirling, se  $n = \sum_{i=1}^r z_i$  e  $z_1, \dots, z_r$  são reais positivos



suficientemente grandes ( $z! = \Gamma(z+1)$ ):

$$p(z_1, \dots, z_r) \geq n^{-r/2} \prod_{i=1}^r \left(\frac{np_i}{z_i}\right)^{z_i} \stackrel{\text{def}}{=} q(z_1, \dots, z_r) \quad (1.9)$$

Tomando-se  $z_i = np_i e^{sx_i / \sum_j p_j e^{sx_j}}$  tem-se

$$(i) \quad \sum_{i=1}^r z_i = n$$

$$(ii) \quad \sum_{i=1}^r z_i x_i = 0$$

$$(iii) \quad q(z_1, \dots, z_r) = n^{-r/2} a^n$$

(Na verdade, fica como Exercício verificar que, tais  $z_1, \dots, z_r$  maximizam  $q(\cdot)$  sob as restrições (i) e (ii).)

No entanto,  $z_1, \dots, z_r$  não são necessariamente inteiros. Sem perda de generalidade, suponhamos  $x_1 \leq x_j$  para  $2 \leq j \leq r$ , e coloquemos:

$$n_j = [z_j] \quad \text{se } j = 2, \dots, r \quad ([z] = \text{parte inteira de } z)$$

$$n_1 = n - \sum_{i=2}^r n_i = z_1 + \sum_{j=2}^r (z_j - [z_j]).$$

Então

$$\sum_{i=1}^r n_i = n$$

$$\sum_{j=1}^r n_j x_j \leq 0$$

e ademais não é difícil ver que existe  $c > 0$  e  $N \in \mathbb{N}$  t.q.

se  $n \geq N$

$$q(n_1, \dots, n_r) \geq cq(z_1, \dots, z_r) = cn^{-r/2} \alpha^n \quad (1.10)$$

o que implicará a afirmação feita. Com efeito, fixando-se  $a > 0$  e intervalo limitado  $I$  existe  $c_1 > 0$  tal que

$$\left( \frac{v'}{av+b} \right)^{av+b} \geq c_1 \left( \frac{v}{av} \right)^{av} \quad (1.11)$$

para todo  $v \geq 1$ , com  $av+b \geq 1$ . (Ver Azencott (1980), Chernoff (1952)).

Conclusão da prova do Teorema 1.2. Queremos provar (1.5) com  $a = 0$  e como notamos, podemos supor  $P(X_1 > 0) > 0$  e  $P(X_1 < 0) > 0$ . Consideremos o seguinte:

1º caso -  $X_1$  v.a. discreta i.e. existem  $x_1, x_2, \dots$  com  $p_i = \mu(x_i) > 0$  e  $\sum_{i=1}^{\infty} p_i = 1$ . Por hipótese existe  $r \in \mathbb{N}$  tal que

$$\min_{1 \leq j \leq r} x_j < 0 < \max_{1 \leq j \leq r} x_j.$$

Se  $\varphi_k(t) = \sum_{i=1}^k p_i e^{tx_i}$ , então  $\varphi_1(t) \leq \varphi_2(t) \leq \dots \nearrow \hat{\mu}(t)$

para todo  $t$ . Pelo Teorema de Dini segue-se que se  $K \subseteq \{t: \mu(t) < +\infty\}$  for um compacto então  $\varphi_k \xrightarrow[k \rightarrow \infty]{} \hat{\mu}$  uniformemente em  $K$ . Daí deduzimos facilmente

$$\inf_t \varphi_k(t) \xrightarrow[k \rightarrow \infty]{} \inf_t \hat{\mu}(t) = e^{-\lambda_{\mu}(0)} \quad (1.12)$$

já que para  $k \geq r$   $\varphi_k(t) \rightarrow +\infty$  quando  $|t| \rightarrow \infty$ . (Exercício 1.5). Temos então  $\log(\inf_t \varphi_k(t)) \xrightarrow{k \rightarrow +\infty} -\lambda_\mu(0)$ . Mas, dado  $k \geq r$  pelo Lema 1.3 tomamos  $c_k > 0$  e  $N_k \in \mathbb{N}$  t.q. se  $n \geq N_k$  e  $n_1, \dots, n_k$  são dados pelo Lema 1.3, temos:

$$P(\bar{X}_n \leq 0) \geq p(n_1, \dots, n_k) \geq c_k n^{-k/2} (\inf \varphi_k(t))^n.$$

Temos então

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\bar{X}_n \leq 0) \geq \log(\inf \varphi_k(t))$$

para todo  $k$ . Fazendo-se  $k \rightarrow +\infty$  segue-se (b) no caso  $X_1$  discreta ( $a = 0$ ).

Caso geral. Para cada  $s \in \mathbb{N}$  definimos

$$X_k^{(s)} = \sum_{i \in \mathbb{Z}} \frac{i}{s} I \left[ X_k \in \left( \frac{i-1}{s}, \frac{i}{s} \right] \right], \quad k \geq 1.$$

Sejam ainda  $\bar{X}_n^{(s)} = \frac{1}{n} \sum_{k=1}^n X_k^{(s)}$  e  $\mu_s$  a lei de  $X_1^{(s)}$ .

Logo,

$$X_k \leq X_k^{(s)} \leq X_k + 1/s \tag{1.13}$$

e temos

$$\hat{\mu}(t) \leq \hat{\mu}_s(t) \leq e^{t/s} \hat{\mu}(t) \quad \text{se } t \geq 0$$

$$e^{t/s} \hat{\mu}(t) \leq \hat{\mu}_s(t) \leq \hat{\mu}(t) \quad \text{se } t \leq 0$$

de modo que  $\hat{\mu}_s(t) \geq e^{-|t|/s} \mu(t)$ , para todo  $t \in \mathbb{R}$ . Aplican-

do-se o Teorema de Dini às funções  $e^{-|t|/s} \hat{\mu}(t)$  e como existem  $A, B$  positivos tais que  $\hat{\mu}(t) \geq A e^{B|t|}$  podemos concluir que dado  $\epsilon > 0$  existe  $s$  t.q.

$$\inf_t \hat{\mu}_s(t) \geq (1-\epsilon) \cdot \inf_t \hat{\mu}(t)$$

e então

$$-\lambda_{\mu_s}(0) \geq -\lambda_{\mu}(0) + \log(1-\epsilon)$$

de modo que por (1.13)

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \log P(X_n \leq 0) &\geq \lim_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n^{(s)} \leq 0) \geq \\ &\geq -\lambda_{\mu}(0) + \log(1-\epsilon). \end{aligned}$$

Fazendo  $\epsilon$  tender a zero obtemos a prova. ■

### Observações:

1. O Teorema de Cramèr-Chernoff fornece uma definição natural de eficiência assintótica para testes da forma anteriormente descrita (exemplificada com o caso Bernoulli). Para esta aplicação referimos ao artigo de Chernoff (1952). Outras aplicações podem ser encontradas nas referências mencionadas no final do §1.1, incluindo também extensões deste resultado.
2. Os exercícios 1.3 a 1.8 contêm exemplos de cálculo da transformada de Cramèr para algumas distribuições em  $\mathbb{R}$ , bem co-

mo algumas propriedades básicas e são, neste ponto, aconselhados.

3. Suponhamos  $E|X_1| < +\infty$ . O Teorema 1.2 dá a existência de  $\lim_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n \in A)$  quando  $A$  é uma semi reta tipo  $(-\infty, x]$  ou  $[x, +\infty)$  e fornece o cálculo de tal limite. Com efeito, é suficiente tomarmos  $A = (-\infty, x]$  para certo  $x \in \mathbb{R}$ . Se  $x \leq EX_1$ , (1.3) e (1.5) implicam que  $\frac{1}{n} \log P(\bar{X}_n \leq x) \rightarrow -\lambda_{\mu}(x)$ . Se  $EX_1 < x$ , pela lei dos grandes números,  $P(\bar{X}_n \leq x) \rightarrow 1$  e então trivialmente  $\frac{1}{n} \log P(\bar{X}_n \leq x) \rightarrow 0$ . Na seção seguinte veremos uma formulação bem mais ampla deste fato.
4. Pela observação anterior o Teorema 1.2 só tem utilidade quando  $\lambda_{\mu}$  não é identicamente nula. Caso contrário seriam necessários resultados mais refinados.

### §1.2. Extensões do Teorema de Cramèr - Chernoff

Com relação ao Teorema de Cramèr - Chernoff, há dois tipos de extensões importantes, e que aparecem de modo natural: queremos permitir  $X_1, X_2, \dots$  tomando valores em  $\mathbb{R}^d$ , ou mesmo em espaços com dimensão infinita; por outro lado também queremos tratar de variáveis com alguma interação entre si, e.g.  $(X_n, n \geq 1)$  Cadeia de Markov ergódica, etc. . Na verdade uma generalização do Teorema 1.2 para  $\mathbb{R}^d$  aparece em Lanford (1961) como uma digressão das idéias lá desenvolvidas para modelos da mecânica esta

tística, envolvendo um grande número de subsistemas que interagem entre si. Neste trabalho ficou evidenciado o papel do funcional "entropia" e idéias básicas nele contidas foram fundamentais para a posterior formalização dos resultados de grandes desvios para vetores aleatórios i.i.d. a valores em espaços mais gerais, cf. Bahadur e Zabell (1979). Um caso particular, e de grande importância na análise de métodos assintóticos em estatística já havia sido tratado por Sanov (1961). Sanov considerou  $Y_1, Y_2, \dots$  vetores aleatórios i.i.d. a valores em  $\mathbb{R}^d$  e estudou problemas de grandes desvios para a distribuição empírica de  $Y_1, \dots, Y_n$ . Se  $\delta_x$  denota a medida de Dirac em  $x$ , então esta distribuição empírica pode ser escrita como  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_j$ , onde  $X_j = \delta_{Y_j}$  para todo  $j \geq 1$ . Trata-se pois de uma extensão do Teorema 1.2, sendo que  $X_1, X_2, \dots$  tomam agora valores no espaço das medidas de probabilidade em  $\mathbb{R}^d$ . Este resultado será enunciado na próxima secção, utilizando-se a formalização proposta por Donsker e Varadhan: para uma classe de borelianos  $A$  no espaço das probabilidades em  $\mathbb{R}^d$  teremos

$$\frac{1}{n} \log P(\bar{X}_n \in A) \rightarrow \inf_{\nu \in A} I_{\pi}(\nu)$$

onde  $\pi$  é a lei comum dos  $Y_n$ , e

$$I_{\pi}(\nu) = \int g \log g d\pi \quad \text{se } \nu \ll \pi \quad \text{e} \quad d\nu = g d\pi$$

$= +\infty$  se  $\nu$  não é absolutamente contínua em relação a  $\pi$ .

Estudo semelhante sobre medidas empíricas

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_j}$ ,  $n \geq 1$ , tem interesse bem mais amplo do que no caso i.i.d.. Um caso interessante é quando  $(Y_n)_{n \geq 1}$  forma uma cadeia de Markov ergódica, de modo que com probabilidade um  $\bar{X}_n \rightarrow \mu$  (fracamente), onde  $\mu$  é a probabilidade estacionária. Neste caso trata-se do problema de grandes desvios com relação ao Teorema Ergódico para cadeias de Markov. Neste tópico uma referência básica é Donsker-Varadhan (1975), que trata de problemas bem mais gerais.

No Capítulo 3 estudamos problemas que envolvem a Teoria de Grandes Desvios para processos de difusão obtidos pela adição de "pequeno" ruído aleatório a sistemas determinísticos. Esta teoria, que estudaremos no Capítulo 2, foi desenvolvida por Freidlin e Wentzell (1970). A extensão do Teorema de Cramér-Chernoff para espaços de Hilbert e de Banach permitiria um outro tratamento para os problemas do Capítulo 2, que de certa forma unificaria os resultados. (Cf. Azencott (1980).) No entanto, razões "pedagógicas" nos levam a preferir o método original de Freidlin e Wentzell em todo o Capítulo 2.

Sejam  $X_1, X_2, \dots$  v.a. i.i.d. a valores em  $\mathbb{R}^d$ , sendo  $\mu$  a lei comum. Uma primeira observação, simples, porém básica e crucial refere-se ao comportamento de  $P(\bar{X}_n \in A)$ , onde  $A$  é um conjunto convexo. Neste caso  $f(n) \stackrel{\text{def}}{=} \log P(\bar{X}_n \in A)$  é superaditiva, o que garante a convergência de  $f(n)/n$ , quando  $n \rightarrow +\infty$ . Isto permite que se obtenham facilmente uma cota inferior para  $\liminf \frac{1}{n} \log P(\bar{X}_n \in A)$ , quando  $A$  é aberto, e uma cota superior

quando  $A$  é compacto, em termos dos limites correspondentes ao caso de conjuntos convexos, e nos leva naturalmente a uma outra definição de  $\lambda_{\mu}(\cdot)$  generalizando-se o Teorema 1.2. Para que se tenha um modo de calcular tal  $\lambda_{\mu}(\cdot)$  o resultado fundamental é que a generalização de (1.2) nos fornece exatamente a mesma função, desde que  $E|X_1| < \infty$ . Esta metodologia é ponto básico dos resultados de Bahadur-Zabell para espaços de Banach, ou mesmo em situações mais gerais (cf. também Donsker-Varadhan (1975)), inspirados parcialmente em idéias de Lanford (1961). Examinaremos a seguir o caso de vetores aleatórios tomando valores em  $\mathbb{R}^d$ , a fim de evitar technicalidades. Entretanto é útil examinarmos a generalidade dos argumentos, já que isto fornecerá exemplos interessantes, tais como o Teorema de Sanov.

Lema 1.4. Seja  $C \subseteq \mathbb{R}^d$  um conjunto convexo (e portanto Boreliano). Então:

$$\lim_n \frac{1}{n} \log P(\bar{X}_n \in C) \stackrel{\text{def}}{=} \iota(C) \text{ existe em } [-\infty, 0],$$

e coincide com  $\sup \frac{1}{n} \log P(\bar{X}_n \in C)$ .

Prova: Fixemos um tal  $C$  convexo; seja  $f(n) = \log P(\bar{X}_n \in C) \stackrel{\text{def}}{=} f(n)$ .

Note que

$$\begin{aligned} P(\bar{X}_n \in C) P(\bar{X}_m \in C) &= P\left(\frac{X_1 + \dots + X_n}{n} \in C\right) P\left(\frac{X_{n+1} + \dots + X_{n+m}}{m} \in C\right) \\ &= P\left(\frac{X_1 + \dots + X_n}{n} \in C, \frac{X_{n+1} + \dots + X_{n+m}}{m} \in C\right) \\ &\leq P(\bar{X}_{n+m} \in C), \end{aligned}$$



pois  $C$  é convexo e  $X_{n+m} = \frac{n}{n+m} \left( \frac{X_1 + \dots + X_n}{n} \right) + \frac{m}{n+m} \left( \frac{X_{n+1} + \dots + X_{n+m}}{m} \right)$ .

Temos pois  $f(n+m) \geq f(n) + f(m)$  e isto implica (ver eg. Billingsley (1965)) que  $\frac{f(n)}{n} \xrightarrow{n \rightarrow \infty} \sup_m \frac{f(m)}{m}$ .

(Obs. Se  $f(n) > -\infty$  para certo  $n$  então  $f(m) > -\infty \quad \forall m \geq n$  e então o limite é finito.) ■

Podemos então fazer a seguinte

Definição 1.5. Seja  $\mu$  uma probabilidade nos borelianos de  $\mathbb{R}^d$ .

Definimos  $\lambda_\mu: \mathbb{R}^d \rightarrow [0, +\infty]$  via

$$\begin{aligned} \lambda_\mu(x) &= -\inf\{\iota(C): C \text{ aberto, convexo, } x \in C\} \\ &= \sup\{-\iota(C): C \text{ aberto, convexo, } x \in C\} \end{aligned} \tag{1.14}$$

É fácil constatar o seguinte:

Lema 1.6.  $\lambda_\mu(\cdot)$  é convexa e semicontínua inferiormente.

Prova: (a) Semicontinuidade inferior. Seja  $x \in \mathbb{R}^d$  e  $a < \lambda_\mu(x)$ .

Podemos então encontrar  $C$ , vizinhança aberta e convexa de  $x$ , tal que  $-\iota(C) > a$ . Ora, se  $y \in C$  por (1.14) também temos  $-\iota(C) \leq \lambda_\mu(y)$ . Isto prova  $\lambda_\mu(x) \leq \lim_{y \rightarrow x} \lambda_\mu(y)$  i.e. a semicontinuidade inferior.

(b) Convexidade. Como  $\lambda_\mu(\cdot)$  é semicontínua, basta verificarmos que

$$\lambda_{\mu} \left( \frac{x+y}{2} \right) \leq \frac{\lambda_{\mu}(x) + \lambda_{\mu}(y)}{2} \quad (1.15)$$

para quaisquer  $x, y \in \mathbb{R}^d$ . Para isto basta provar que se  $C$  é um aberto, convexo tal que  $\frac{x+y}{2} \in C$  então existem  $C_1, C_2$  abertos convexos tais que  $x \in C_1, y \in C_2$  e tais que

$$-t(C) \leq -\frac{1}{2} (t(C_1) + t(C_2)) \quad (1.16)$$

De fato, isto implica (por (1.14))

$$-t(C) \leq \frac{1}{2} (\lambda_{\mu}(x) + \lambda_{\mu}(y))$$

para todo conjunto  $C$  aberto convexo t.q.  $\frac{x+y}{2} \in C$ , e (1.15) decorre imediatamente, pela definição (1.14).

Ora dado  $C$  cf. acima podemos tomar  $C_1, C_2$  abertos convexos tais que  $x \in C_1, y \in C_2$  e

$$\frac{1}{2} C_1 + \frac{1}{2} C_2 \subseteq C \quad (1.17)$$

de modo que

$$\left[ \frac{X_1 + \dots + X_n}{n} \in C_1 \right] \cap \left[ \frac{X_{n+1} + \dots + X_{2n}}{n} \in C_2 \right] \subseteq \left[ \bar{X}_{2n} \in C \right].$$

Isto implica

$$P(\bar{X}_{2n} \in C) \geq P(\bar{X}_n \in C_1) P(\bar{X}_n \in C_2)$$

e portanto

$$\frac{1}{2n} \log P(\bar{X}_{2n} \in C) \geq \frac{1}{2} \left( \frac{1}{n} \log P(\bar{X}_n \in C_1) + \frac{1}{n} \log P(\bar{X}_n \in C_2) \right).$$

Fazendo  $n \rightarrow \infty$  tem-se (1.16) e a prova do lema. ■

Para a extensão do Teorema 1.2 é natural colocarmos a seguinte:

Definição 1.7. Sendo  $A \in \mathcal{B}(\mathbb{R}^d) \stackrel{\text{def}}{=} \{B \subset \mathbb{R}^d: B \text{ é conjunto de Borel}\}$ , definimos:

$$\underline{I}(A) = \lim_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n \in A) \quad (1.18a)$$

$$\bar{I}(A) = \overline{\lim}_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n \in A), \quad (1.18b)$$

e quando  $\underline{I}(A) = \bar{I}(A)$  denotamos por  $I(A)$  o valor comum. Seja ainda

$$\Lambda(A) = \inf_{x \in A} \lambda_{\mu}(x). \quad (1.19)$$

Podemos então provar a seguinte

Proposição 1.8. (Bahadur - Zabell (1979), Donsker-Varadhan(1975))

Seja  $A \in \mathcal{B}(\mathbb{R}^d)$ . Tem-se:

- (1)  $-\Lambda(\overset{\circ}{A}) \leq \underline{I}(A)$
- (2) Se  $\bar{A}$  for compacto,  $\bar{I}(A) \leq -\Lambda(\bar{A})$
- (3) Se  $\bar{A}$  for compacto e  $\inf_{x \in \overset{\circ}{A}} \lambda_{\mu}(x) = \inf_{x \in \bar{A}} \lambda_{\mu}(x)$  então

$\iota(A)$  existe e  $\iota(A) = -\Lambda(A)$

(4) Se  $A$  for uma união finita de abertos, convexos, então  $\iota(A)$  existe e  $\iota(A) = -\Lambda(A)$ .

Prova: (1) Como  $\underline{\iota}(A) \leq \underline{\iota}(B)$  se  $A \subseteq B$ , basta provar no caso de  $A$  aberto. Neste caso, dado  $x \in A$  existe  $\delta > 0$  t.q.

$C = B_\delta(x) \subseteq A$ ; onde  $B_\delta(x) = \{y: |y-x| < \delta\}$ ; sendo  $B_\delta(x)$  aberto e convexo obtemos  $-\lambda_\mu(x) \leq \iota(B_\delta(x)) \leq \underline{\iota}(A)$ .

Dáí segue-se (1).

(2) Novamente basta considerarmos o caso  $A$  compacto, pois  $\bar{\iota}(A) \leq \bar{\iota}(\bar{A})$ . Queremos provar que

$$\bar{\iota}(A) \leq \sup_{x \in A} (-\lambda_\mu(x)).$$

Mas, por (1.14), se  $a > \sup_{x \in A} -\lambda_\mu(x)$ , então para cada  $x \in A$  existe  $C_x$  aberto, convexo, e contendo  $x$ , tal que  $\iota(C_x) < a$ . Sendo  $A$  compacto existem  $x_1, \dots, x_m \in A$  (para certo  $m \geq 1$ ) tais que  $A \subseteq \bigcup_{i=1}^m C_{x_i}$ ; deste modo (Exercício 1.11)

$$\bar{\iota}(A) \leq \bar{\iota} \left( \bigcup_{i=1}^m C_{x_i} \right) \leq \max_{i=1, \dots, m} \iota(C_{x_i}) < a$$

e (2) segue.

(3) Decorre imediatamente de (1) e (2)

(4) Para quaisquer  $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R}^d)$ ,  $m \geq 1$  tem-se

$$\bar{t}\left(\bigcup_{i=1}^m B_i\right) \leq \max_{i=1, \dots, m} \bar{t}(B_i) \quad (1.20)$$

$$\underline{t}\left(\bigcup_{i=1}^m B_i\right) \geq \max_{i=1, \dots, m} \underline{t}(B_i) \quad (1.21)$$

como se verifica facilmente (Exercício 1.11). Daí, se

$\underline{t}(B_1), \dots, \underline{t}(B_m)$  existem então também  $\underline{t}\left(\bigcup_{i=1}^m B_i\right)$  existe e vale  $\max_{i=1, \dots, m} \underline{t}(B_i)$ .

Como consequência da observação anterior, se  $A = \bigcup_{i=1}^m C_i$  onde  $C_1, \dots, C_m$  são abertos convexos e  $m \geq 1$  então existe  $\underline{t}(A)$ , e  $\underline{t}(A) = \max_{i=1, \dots, m} \underline{t}(C_i)$ . Por outro lado  $\Lambda(A) = \min_{i=1, \dots, m} \Lambda(C_i)$  e basta então provarmos que  $\underline{t}(C_i) = -\Lambda(C_i)$   $i = 1, \dots, m$ , ou seja podemos supor  $A$  convexo e aberto. Por (1) temos  $-\Lambda(A) \leq \underline{t}(A)$ . Para a desigualdade recíproca, o caso  $\underline{t}(A) = -\infty$  é trivial e portanto suponhamos  $\underline{t}(A) > -\infty$ . Para cada  $\epsilon > 0$ , podemos tomar  $n_\epsilon$  tal que

$$-\epsilon + \underline{t}(A) \leq \frac{1}{n_\epsilon} \log P(\bar{X}_{n_\epsilon} \in A).$$

Como  $\mu(\mathbb{R}) = 1$  (finita) e  $A$  é aberto, convexo não é difícil ver que existe um conjunto compacto e convexo  $K_\epsilon \subset A$  tal que

$$P(\bar{X}_{n_\epsilon} \in K_\epsilon) \geq (1-\epsilon) P(\bar{X}_{n_\epsilon} \in A),$$

e então

$$\begin{aligned} \log P(X_{n_\epsilon} \in K_\epsilon) &\geq \log P(\bar{X}_{n_\epsilon} \in A) + \log(1 - \epsilon) \geq \\ &\geq n_\epsilon(\iota(A) - \epsilon) + \log(1 - \epsilon) \end{aligned}$$

Ademais, se  $f(n) \stackrel{\text{def}}{=} \log P(\bar{X}_n \in K_\epsilon)$ , sabemos que  $f(m+n) \geq f(m) + f(n)$  pois  $K_\epsilon$  é convexo. Daí:

$$f(kn_\epsilon) \geq kn_\epsilon(\iota(A) - \epsilon) + k \log(1 - \epsilon) \quad \text{para todo } k \geq 1$$

e

$$\iota(K_\epsilon) = \sup_k \frac{f(kn_\epsilon)}{kn_\epsilon} \geq \iota(A) - \left( \epsilon + \frac{|\log(1-\epsilon)|}{n_\epsilon} \right)$$

como  $\log(1-\epsilon) \rightarrow 0$  se  $\epsilon \rightarrow 0$ , e para cada  $\epsilon > 0$  (por (2))

$$-\Lambda(A) \geq -\Lambda(K_\epsilon) \geq \iota(K_\epsilon)$$

obtemos  $-\Lambda(A) \geq \iota(A)$ , concluindo a prova. ■

Para que a Proposição 1.8 possa ser vista como uma extensão do Teorema 1.2 dois pontos precisam ser considerados:

- (i) Quando podemos eliminar a hipótese de compacidade em (2) na Proposição 1.8?
- (ii) Como "calcular"  $\lambda_\mu(\cdot)$ ?

A proposição abaixo mostra que no caso  $E|X_1| < +\infty$  (em  $\mathbb{R}^d$ )  $\lambda_\mu(\cdot)$  pode ser obtida pela extensão natural de (1.2)  $\mathbb{R}^d$ .  
(Notação: se  $x, y \in \mathbb{R}^d$ ,  $\langle x, y \rangle$  denota o produto interno usual.)

Proposição 1.9. Seja  $\mu$  uma probabilidade em  $\mathbb{R}^d$ . Definimos a transformada de Laplace  $\hat{\mu}: \mathbb{R}^d \rightarrow (0, +\infty]$  via

$$\hat{\mu}(t) = \int_{\mathbb{R}^d} e^{\langle t, x \rangle} \mu(dx), \quad t \in \mathbb{R}^d. \quad (1.22)$$

Se  $\int |x| \mu(dx) < +\infty$ , temos então

$$\lambda_{\mu}(x) = \sup_{t \in \mathbb{R}^d} (\langle t, x \rangle - \log \hat{\mu}(t)) \quad (1.23)$$

Prova: caso d=1. Sem perda de generalidade supomos  $EX_1 = 0$ .

Seja

$$\tilde{\lambda}(x) \stackrel{\text{def}}{=} \sup(tx - \log \hat{\mu}(t)).$$

Então: (a)  $\tilde{\lambda}, \lambda$  são ambas semicontínuas inferiormente, convexas, com valores em  $[0, +\infty]$ .

(b)  $\tilde{\lambda}(0) = 0 = \lambda(0)$ . (Com efeito, pela Lei dos Grandes Números  $\lambda(-\epsilon, \epsilon) = 0$  para cada  $\epsilon > 0$  e portanto  $\lambda(0) = 0$ . Por outro lado  $\tilde{\lambda}(0) = -\inf(\log \hat{\mu}(t))$  e aplicando a desigualdade de Jensen obtemos

$$\log \hat{\mu}(t) \geq 0 = \log \hat{\mu}(0), \quad \text{i.e.,} \quad \tilde{\lambda}(0) = 0).$$

As observações (a) e (b) acima implicam a existência de  $p, q \in [0, +\infty)$  tais que  $\lambda(\tilde{\lambda})$  é contínua em  $[0, p)$  ( $[0, q)$ , resp.), vale  $+\infty$  em  $(p, +\infty)$  ( $(q, +\infty)$  resp.) e é contínua à esquerda em  $p$  ( $q$  resp.).

(Notação:  $\lambda(y^-)$ ,  $\lambda(y^+)$  denotam como é usual os limites à esquerda e à direita, resp. quando tais existem.)

Ademais, o Teorema 1.2(a) implica que para cada  $x > 0$ , e  $\epsilon > 0$  suficientemente pequeno

$$\iota(x-\epsilon, x+\epsilon) \leq \iota(x-\epsilon, +\infty) \leq -\tilde{\lambda}(x-\epsilon)$$

de modo que

$$-\lambda(x) = \lim_{\epsilon \downarrow 0} \iota(x-\epsilon, x+\epsilon) \leq -\tilde{\lambda}(x^-) \tag{1.24a}$$

Por outro lado, pelo Teorema 1.2(b) e pela Proposição 1.8 (ítem (3)) temos: ( $x \geq 0$ )

$$\begin{aligned} -\tilde{\lambda}(x) &\leq \iota[x, +\infty) \leq \iota[x] \vee \iota(x, +\infty) \leq \\ &\leq -\lambda(x) \vee -\Lambda(x, +\infty) . \end{aligned} \tag{1.24b}$$

Ora se  $x \geq 0$  temos então

$$\Lambda(x, +\infty) = \inf_{y > x} \lambda_{\mu}(y) = \lim_{y \downarrow x} \lambda_{\mu}(y) \stackrel{\text{det}}{=} \lambda_{\mu}(x^+)$$

pois  $\lambda_{\mu}(\cdot)$  é s.c.i., convexa, e  $0 = \lambda_{\mu}(0) \leq \lambda_{\mu}(y)$  para todo  $y$ . Conseqüentemente temos, por (1.24):

$$\tilde{\lambda}(x^-) \leq \lambda(x) \quad \text{se } x > 0$$

$$\lambda(x) \wedge \lambda(x^+) \leq \tilde{\lambda}(x) \quad \text{se } x \geq 0$$

Com isto é fácil concluir que  $p = q$  e então também  $\lambda(x) = \tilde{\lambda}(x)$  para todo  $x \geq 0$ . Raciocínio análogo aplica-se a



$x \leq 0$ , provando (1.23) no caso  $d = 1$ .

Caso geral ( $d \geq 1$ ). A observação principal, sugerindo uma redução a caso unidimensional é a de que se  $\mu$  é probabilidade em  $\mathbb{R}^d$ , e  $\lambda_\mu$  é definida por (1.14), então

$$\begin{aligned} \lambda_\mu(x) &= -\inf\{\iota(H) : H \text{ semiespaço aberto, } x \in H\} \\ &= \sup\{-\iota(H) : H \text{ semiespaço aberto, } x \in H\} \end{aligned} \quad (1.25)$$

Com efeito, pela definição

$$\lambda_\mu(x) \geq \sup\{-\iota(H) : H \text{ semiespaço aberto, } x \in H\}$$

pois um semiespaço é um conjunto convexo. Para a recíproca podemos supor  $\lambda_\mu(x) > 0$ . Seja  $\epsilon > 0$  e mostremos que existe  $H$  semiespaço aberto com  $x \in H$  e tal que  $\lambda_\mu(x) - \epsilon \leq -\iota(H)$ . Mas o conjunto  $C = \{y : \lambda_\mu(y) \leq \lambda_\mu(x) - \epsilon\}$  é convexo e fechado pois  $\lambda_\mu(\cdot)$  é convexa e semicontínua inferiormente. Ademais  $x \notin C$ . Ora em  $\mathbb{R}^d$  isto implica a existência de um hiperplano que separa (estritamente)  $x$  e  $C$ , i.e., existe  $y \in \mathbb{R}^d$  tal que

$$\langle x, y \rangle > \sup_{c \in C} \langle c, y \rangle = m$$

e portanto se definimos o semiespaço aberto

$$H = \{z : \langle z, y \rangle > m\}$$

temos que  $x \in H$  e  $H \cap C = \emptyset$ , e pela Proposição 1.8 (3) temos

$$-\iota(H) = \Lambda(H) = \inf_{z \in H} \lambda_{\mu}(z) \geq \lambda_{\mu}(x) - \epsilon,$$

o que prova (1.25).

Com (1.25) a redução ao caso unidimensional é bastante simples. Note antes de tudo, que todo semiespaço aberto que contém  $x \in \mathbb{R}^d$  é da forma

$$H(t,r) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^d: \langle t,y \rangle < r\}$$

para certo  $t \in \mathbb{R}^d$  e  $r \in \mathbb{R}$  de modo que  $\langle t,x \rangle < r$ .

Para cada  $t \in \mathbb{R}^d$  consideramos o funcional  $\langle t, \cdot \rangle: \mathbb{R}^d \rightarrow \mathbb{R}$  e seja  $\mu_t$  a medida induzida em  $\mathbb{R}$  pela medida  $\mu$ . Definimos as variáveis aleatórias  $\langle t, X_n \rangle$ ,  $n \geq 1$ , que serão i.i.d. com lei  $\mu_t$ . Seja ainda  $\lambda_t(\cdot) \stackrel{\text{def}}{=} \lambda_{\mu_t}(\cdot)$  a transformada de Cramèr correspondente (em  $\mathbb{R}$ ). Se

$$\iota_t(A) = \lim_{n \rightarrow +\infty} \frac{1}{n} \log P(\langle t, \bar{X}_n \rangle \in A), \text{ para } A \text{ intervalo } \subset \mathbb{R},$$

então, por (1.25) no caso  $d = 1$ , se  $s \in \mathbb{R}$ :

$$\lambda_t(s) = \max\left\{\sup_{u > s} -\iota_t(-\infty, u), \sup_{u < s} -\iota_t(u, +\infty)\right\}$$

e portanto

$$\lambda_t(\langle t, x \rangle) = \max\left\{\sup_{r > \langle t, x \rangle} -\iota(H(t,r)), \sup_{r < \langle t, x \rangle} -\iota(H(-t, -r))\right\},$$

de modo que novamente por (1.25) aplicada a  $\mu$  obtemos:

$$\lambda_{\mu}(x) = \sup_{t \in \mathbb{R}} \lambda_t(\langle t, x \rangle)$$

e então pelo caso  $d = 1$

$$\begin{aligned} \lambda_{\mu}(x) &= \sup_{t \in \mathbb{R}^d} \lambda_t(\langle t, x \rangle) \\ &= \sup_{t \in \mathbb{R}^d} \sup_{s \in \mathbb{R}} (s \langle t, x \rangle - \log \hat{\mu}_t(s)) \\ &= \sup_{t \in \mathbb{R}^d} \sup_{s \in \mathbb{R}} (s \langle st, x \rangle - \log \hat{\mu}(st)) \\ &= \sup_{t \in \mathbb{R}^d} (\langle t, x \rangle - \log \hat{\mu}(t)), \end{aligned}$$

concluindo a prova. ■

Observações com vistas à extensão dos resultados anteriores para espaços infinito-dimensionais.

Nossas discussões anteriores são válidas em um contexto bem mais geral. Na verdade, o método acima descrito é essencialmente o que Bahadur e Zabell desenvolveram para tratar de grandes desvios para a média amostral de uma seqüência de vetores aleatórios i.i.d., com valores em espaços possivelmente infinito-dimensionais. (na seção seguinte veremos uma importante aplicação, o Teorema de Sanov).

Consideremos  $X_1, X_2, \dots$  vetores aleatórios independentes e identicamente distribuídos a valores em um certo espaço vetorial topológico  $E$ , munido de sua  $\sigma$ -álgebra de Borel  $\mathcal{B}(E)$ .

Seja  $\mu$  a probabilidade em  $\mathcal{B}(E)$  que é a lei comum. Para que a própria questão faça sentido precisamos antes de tudo que  $\bar{X}_n$  seja vetor aleatório (para cada  $n \geq 1$ ). Também, o método descrito acima baseia-se na "abundância de conjuntos convexos. Isto sugere as seguintes hipóteses sobre  $E$  e  $\mu$ : (Cf. Azencott (1980) e Bahadur-Zabell (1979), onde a situação é um pouco mais geral.)

(c1)  $E$  espaço vetorial topológico, localmente convexo e de Hausdorff, sendo  $\mathcal{B}(E)$  a sua  $\sigma$ -álgebra de Borel.  $\mu$  probabilidade em  $(E, \mathcal{B}(E))$ .  $X_1, X_2, \dots$  variáveis aleatórias i.i.d.  $X_n: \Omega \rightarrow E$  onde supomos que o espaço de probabilidade  $(\Omega, \mathcal{G}, P)$  seja completo.

(c2) Existe  $C$  convexo, fechado,  $C \subseteq E$  de modo que  $\mu(C) = 1$  e em  $E$  é possível construir topologia mais fina que a dada em (c1) de modo que com a nova topologia induzida  $C$  seja um espaço polonês (metrizável, completo e separável).

As condições podem parecer complicadas, mas alguns exemplos ajudarão a perceber os casos mais comuns:

Exemplos: (a)  $E = \mathbb{R}^d$  e  $\mu$  probabilidade em  $\mathcal{B}(E)$ . De modo mais geral  $E$  um espaço de Fréchet separável e  $\mu$  qualquer probabilidade em  $\mathcal{B}(E)$ . (Aqui como  $E$  já é polonês em (c2) podemos tomar  $C = E$  e usar a mesma topologia já dada, de modo que o resultado se aplicará a qualquer probabilidade  $\mu$  em  $\mathcal{B}(E)$ .)

Observações: O ex. (a) inclui espaços de Banach(ou Hilbert) separáveis: e.g.  $C[0,1]$ ,  $L^2[0,1]$ , etc.... que são importantes para vários processos estocásticos, principalmente processos Gaussianos.

$$(b) \text{ Seja } E = \mathfrak{M}(\mathbb{R}^d) = \left\{ \eta: \mathfrak{B}(\mathbb{R}^d) \rightarrow \mathbb{R} / \begin{array}{l} \eta \text{ é medida com sinal} \\ \text{e } \|\eta\| < \infty \end{array} \right\}$$

onde  $\|\eta\| = \eta^+(\mathbb{R}^d) + \eta^-(\mathbb{R}^d)$  é a variação total. Consideremos em

$E$ , que é o dual de  $C_b(\mathbb{R}^d) \stackrel{\text{def}}{=} \{f: \mathbb{R}^d \rightarrow \mathbb{R} / f \text{ é contínua, limitada}\}$ , a topologia fraca\* (ao invés da variação total).

Se tivermos  $C = \{\eta \in E: \eta \text{ é medida de probabilidade}\} \stackrel{\text{def}}{=} \mathfrak{M}_1(\mathbb{R}^d)$  então  $C$  é um espaço polonês, quando munido da topologia fraca\*. Por isto (c2) se verifica se  $\mu$  for uma probabilidade tal que  $\mu(C) = 1$ .

Nestas notas não queremos aprofundar muito as questões mais técnicas. Por isto, apenas uns comentários sobre o porquê destas hipóteses:

(1) (c2) garante que se  $X, Y$  são variáveis aleatórias a valores em  $C$  então  $X+Y$  também é variável aleatória. Daí, se  $X_1, \dots, X_n$  são variáveis aleatórias a valores em  $C$ ,  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  também o é.

(2) O fato de supor  $(\Omega, \mathcal{G}, P)$  completo não impõe restrição alguma, como sabemos. Se preferir, suponha que as v.a.  $X_1, X_2, \dots$

assumem valores em  $C$  e então pode omitir esta technicalidade.

(3) As condições (c1), (c2) permitem a reprodução das "mesmas" demonstrações utilizadas anteriormente. Na verdade o método foi proposto por Bahadur e Zabell justamente para tratar o caso geral; fizemos  $E = \mathbb{R}^d$  (inicialmente) para evitar que o leitor se preocupasse com aspectos mais técnicos.

Inicialmente notemos que sob (c1), (c2), o Lema 1.4 vale de forma quase idêntica, bastando acrescentar a hipótese  $C \in \mathcal{B}(E)$ , antes desnecessária. Da mesma forma, definindo-se a transformada de Cramèr  $\lambda_\mu(\cdot)$  por (1.14), o Lema 1.6 continua válido neste contexto. Observe que aí utilizamos realmente a hipótese de convexidade local (cada ponto possui sistema fundamental de vizinhanças convexas). Como antes, definimos  $\underline{t}(A)$  e  $\bar{t}(A)$  para  $A \in \mathcal{B}(E)$ , através de (1.18) e  $\Lambda(A)$  é definido por (1.19). Com isto, a Proposição 1.8 continua válida: de fato, para o ítem (1) basta utilizarmos a convexidade local para obtermos o conjunto  $C$  que lá tomamos como uma bola. O ítem (2) é idêntico. No ítem (4) precisamos garantir que dados  $\epsilon, n$  existe  $K_\epsilon$  compacto convexo, contido em  $A$  tal que

$$P(\bar{X}_n \in K_\epsilon) \geq (1-\epsilon) P(\bar{X}_n \in A).$$

Isto vem do seguinte lema de Bahadur e Zabell (1979).

Lema 1.10. Suponha que  $E, \mu$  verifiquem (c2). Então para todo  $A$  aberto, convexo de  $E$  e todo  $\epsilon > 0$  existe  $K$  compacto convexo,  $K \subseteq A$  tal que  $\mu(A \setminus K) < \epsilon$ . (Omitiremos a prova).

cf. Azencott (1980) ou Bahadur - Zabell (1979).)

Com relação à caracterização de  $\lambda_{\mu}(\cdot)$  dada pela Proposição 1.9, o ponto importante consiste exatamente na extensão de (1.25). Mas isto decorre da possibilidade de separação (estrita) através de hiperplano de um conjunto convexo fechado  $C$  e um ponto  $x \notin C$ , e sob (c1) e (c2) isto é possível.

Antes de enunciarmos a extensão da Proposição 1.9, façamos algumas definições.

Definição 1.11. Sejam  $E, \mu$  nas condições (c1), (c2).  $E'$  denota o dual topológico de  $E$  i.e.

$$E' = \{t: E \rightarrow \mathbb{R} / t \text{ é funcional linear, contínuo}\}.$$

(Para  $t \in E'$ ,  $x \in E$  o valor  $t(x)$  será denotado por  $\langle t, x \rangle$ , como é usual, e  $w^*$  denota a topologia fraca\* em  $E'$ .)

Podemos definir a transformada de Laplace  $\hat{\mu}: E' \rightarrow (0, +\infty]$  via

$$\hat{\mu}(t) = \int_E e^{\langle t, x \rangle} \mu(dx), \quad t \in E'.$$

Como antes  $\log \hat{\mu}(t)$  é convexa e semicontínua inferiormente para a topologia  $w^*$  em  $E'$ . Ademais, pelas observações anteriores tem-se a seguinte:

Proposição 1.9'. Se  $E, \mu$  verificam (c1), (c2) e ademais  $\int_E |\langle t, x \rangle| \mu(dx) < +\infty$  para cada  $t \in E'$ , então para todo  $x \in E$

$$\lambda_{\mu}(x) = \inf_{t \in E'} (\langle t, x \rangle - \log \hat{\mu}(t)).$$

Ou seja a hipótese de integrabilidade das variáveis aleatórias  $\langle t, X_1 \rangle$  com  $t \in E'$  garante a caracterização anterior de  $\lambda_{\mu}(\cdot)$ .

A utilização da Proposição 1.9 ( ou Proposição 1.9') nos permite identificarmos  $\lambda_{\mu}(\cdot)$  em muitas situações. Alguns exemplos estão nos Exercícios. Também nos permite identificarmos o caso em que esta teoria não dá informação, i.e. quando  $\lambda_{\mu}(x) \equiv 0$ . Isto ocorre se e só se  $\hat{\mu}(t) = +\infty$  para todo  $t \in E' \setminus \{0\}$ . Veja os exercícios para exemplos neste sentido. Nestes casos é necessário fazer algo mais refinado, pois  $\frac{1}{n} \log P(\bar{X}_n \in A)$  tenderá em geral a zero, e isto não dá quase informação sobre o comportamento de  $P(\bar{X}_n \in A)$  quando  $n \rightarrow +\infty$ .

Há um ponto na Proposição 1.9' que precisaria ser melhorado: a hipótese de compacidade no item (2). Esta hipótese especialmente em espaços de Banach infinito-dimensionais, parece extremamente restritiva. Entretanto sua eliminação baseia-se em decaimento muito rápido das caudas de  $\mu$ , como é o caso do Lema 6.1 de Azencott (1980) supondo a existência, para cada  $R > 0$ , de um compacto  $K_R$  tal que  $P(\bar{X}_n \notin K_R) \leq e^{-nR}$  para  $n$  suficientemente grande. No caso de  $\mathbb{R}^d$  isto pode ser obtido com a hipótese de que  $\hat{\mu}(t) < +\infty$  para todo  $t$  em alguma vizinhança da origem o que implica  $\lambda_{\mu}(x) \rightarrow +\infty$  se  $|x| \rightarrow \infty$ . Alguns exemplos aparecem nos exercícios deste capítulo. Referimos a Dansker-Varadhan (1975) para uma maior discussão sobre a desi-



gualdade  $\bar{\lambda}(A) \leq -\lambda(\bar{A})$ , e o problema relacionado da compacidade de de  $\{x: \lambda_{\mu}(x) \leq b\}$  se  $b < +\infty$ .

### §1.3 Teorema de Sanov. Grandes desvios de distribuições empíricas.

O conteúdo básico desta seção, conforme indicado acima, é um teorema provado inicialmente por Sanov em 1975 e que trata do comportamento limite, quando  $n \rightarrow +\infty$  de

$\log P \left[ \sup_x |F_n(x) - G(x)| < \epsilon \right]$  onde  $\epsilon > 0$ ,  $F_n$  é a distribuição empírica de  $Y_1, \dots, Y_n$  variáveis aleatórias i.i.d., reais, com uma certa distribuição  $F$ , e  $G$  é uma função de distribuição em  $\mathbb{R}$  tal que  $I \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \frac{dF}{dG}(x) G(dx)$  exista. Neste caso ele prova

$$\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow +\infty} \frac{1}{n} \log P(\sup_x |F_n(x) - G(x)| < \epsilon) = I \quad (1.26)$$

Em particular (1.26) forneceu uma interpretação probabilística da entropia de uma distribuição  $F$  absolutamente contínua, em  $[0,1]$ . (cf. Sanov (1961).)

O método desenvolvido por Bahadur-Zabell (1979), e Donsker-Varadhan (1975) que estudamos (parcialmente) na seção 1.2 permite tratarmos a questão como um problema de grandes desvios para a média amostral das variáveis  $\delta_{Y_1}, \dots, \delta_{Y_n}$  que tomam valores no espaço das medidas de probabilidade em  $\mathbb{R}$ . Com isto, também não há razão para nos restringirmos a  $\mathbb{R}$ . Por "sim-

plicidade" suporemos  $Y_1, Y_2, \dots$  tomando valores em  $\mathbb{R}^d$ , mas tudo se aplica igualmente se as v.a.'s  $Y_n$ ,  $n \geq 1$  tomarem valores em um espaço polonês. Os resultados colocados na mesma formulação da Proposição 1.9' podem ser comparados com (1.26). Para isto veja alguns exercícios do Capítulo 2. (Compare com os Teoremas 2.2 e 2.3.)

Definição 1.2 (Kullback)

Sejam  $\pi, \nu$  probabilidades em  $\mathcal{M}(\mathbb{R}^d)$ . A informação de Kullback de  $\nu$  com relação a  $\pi$  é dada por

$$I_{\pi}(\nu) = \int_{\mathbb{R}^d} f \log f \, d\pi, \text{ se } \nu \ll \pi, \text{ sendo } f = \frac{d\nu}{d\pi} \quad (*)$$
$$(1.27)$$

$= +\infty$ , caso contrário.

Logo  $I_{\pi}: \mathcal{M}_1(\mathbb{R}^d) \rightarrow [0, +\infty]$ . (Heuristicamente  $I_{\pi}(\nu)$  grande significa  $\nu$  muito diferente de  $\pi$ .) Consideramos aqui, como em §1.2,  $\mathcal{M}_1(\mathbb{R}^d) = \{P: \mathcal{M}(\mathbb{R}^d) \rightarrow [0, 1] \text{ t.q. } P \text{ é probabilidade } \sigma\text{-aditiva}\}$  com a topologia  $w^*$ , nele induzida como subespaço (topológico) de  $\mathcal{M}(\mathbb{R}^d)$  que é o dual de  $C_b(\mathbb{R}^d) \stackrel{\text{def}}{=} \{f: \mathbb{R}^d \rightarrow \mathbb{R} \mid f \text{ é contínua e limitada}\}$  com a norma do supremo. (Exemplo (b) de espaços verificando (c1), (c2).)

Podemos então enunciar o seguinte

---

(\*) Por definição  $0 \cdot \log 0 = 0$  i.e.  $\int f \log f \, d\pi \Big|_{[f>0]}$

Teorema 1.13 (Sanov (1961), Donsker-Varadhan (1975), Bahadur-Zabell (1979)).

Sejam  $Y_1, Y_2, \dots$  v.a.'s i.i.d. assumindo valores em  $\mathbb{R}^d$ . Sejam ainda  $\pi$  a lei comum desta variáveis e  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$  a lei empírica de  $Y_1, \dots, Y_n$ , para cada  $n \geq 1$ . Então:

(a) Para todo  $A$  boreliano de  $m_1(\mathbb{R}^d)$ :

$$-I_{\pi}(A) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n \in A) \leq \overline{\lim}_{n \rightarrow +\infty} \frac{1}{n} \log P(\bar{X}_n \in A) \leq -I_{\pi}(\bar{A}) \quad (1.28)$$

onde 
$$I_{\pi}(B) = \inf_{\nu \in B} I_{\pi}(\nu).$$

(b) No caso de  $A$  ser união finita de abertos convexos de  $m_1(\mathbb{R}^d)$  tem-se que

$$\frac{1}{n} \log P(\bar{X}_n \in A) \xrightarrow{n \rightarrow +\infty} -I_{\pi}(A).$$

(c) Sendo  $X_n = \delta_{Y_n}$  então  $X_1, X_2, \dots$  são variáveis aleatórias i.i.d. a valores no espaço vetorial topológico  $m(\mathbb{R}^d)$ . Se  $\mu$  denota a lei comum então

$$\begin{aligned} \lambda_{\mu}(\nu) &= I_{\pi}(\nu) \text{ se } \nu \text{ é uma probabilidade} \\ &= +\infty \text{ caso contrário} \end{aligned} \quad (1.29)$$

Também, se  $b < +\infty$ , o conjunto  $\{\nu: \lambda_{\pi}(\nu) \leq b\}$  é compacto, na topologia  $w^*$ .

Esquema de prova (Ver, por exemplo, Azencott (1980).)

Claramente (c) é o ponto básico. Estamos considerando o  $E = \mathbb{M}(\mathbb{R}^d)$  com a topologia  $w^*$ . Se  $X_n \stackrel{\text{def}}{=} \delta_{Y_n}$ ,  $n \geq 1$  então as condições (c1), (c2) se verificam, tomando-se  $C = \mathbb{M}_1(\mathbb{R}^d)$ . Pela Proposição 1.9' e Lema 1.10 devemos provar:

(i) (1.29) é válido;

(ii) dado  $r > 0$  existe  $K_r$  compacto de  $\mathbb{M}_1(\mathbb{R}^d)$  tal que  $P(\bar{X}_n \notin K_r) \leq e^{-nr}$  para  $n$  suficientemente grande.

O item (ii) é mais técnico e referimos a Azencott (1980) ou B-Z (1979), D-V (1980). Quanto a (1.29) vejamos inicialmente que  $\lambda_\mu(v) = +\infty$  quando  $v \notin \mathbb{M}_1(\mathbb{R}^d)$ . De fato neste caso, pela convexidade local existe  $U$  aberto, convexo t.q.  $v \in U \subseteq \mathbb{M}(\mathbb{R}^d) \setminus \mathbb{M}_1(\mathbb{R}^d)$ . Mas então  $P(\bar{Y}_n \in U) = 0$  para todo  $n$  e obtemos  $l(U) = -\infty$  de modo que  $\lambda_\mu(v) \geq l(U) = +\infty$ . Seja agora  $v \in \mathbb{M}_1(\mathbb{R}^d)$ . Da Proposição 1.9, estendida às condições (c1), (c2), temos:

$$\lambda_\mu(v) = \sup_{f \in C_b(\mathbb{R}^d)} \left[ \langle v, f \rangle - \log \hat{\mu}(f) \right]$$

sendo  $\langle v, f \rangle = \int_{\mathbb{R}^d} f(x) v(dx)$ . Como  $\mu$  é a lei de  $\delta_{Y_1}$

$$\hat{\mu}(f) = \int_{\mathbb{M}_1(\mathbb{R}^d)} e^{\langle f, \eta \rangle} \mu(d\eta) = \int_{\mathbb{R}^d} e^{f(x)} \pi(dx)$$

e temos

$$\lambda_{\mu}(\nu) = \sup_{f \in C_b(\mathbb{R}^d)} \left[ \int_{\mathbb{R}^d} f d\nu - \log \int_{\mathbb{R}^d} e^f d\pi \right]$$

Para verificarmos que  $\lambda_{\mu}(\nu) \leq I_{\pi}(\nu)$  basta supor  $I_{\pi}(\nu) < \infty$ . Mas então  $d\nu = h d\pi$  para certa  $h \geq 0$ , e se  $f \in C_b(\mathbb{R}^d)$

$$\int f d\nu - I_{\pi}(\nu) = \int (f - \log h) d\nu$$

e pela desigualdade de Jensen

$$\exp\left(\int f d\nu - I_{\pi}(\nu)\right) \leq \int e^f d\pi$$

de modo que

$$\int f d\nu - \log \int e^f d\pi \leq I_{\pi}(\nu),$$

provando  $\lambda_{\mu}(\nu) \leq I_{\pi}(\nu)$ . Para a recíproca podemos supor  $\lambda_{\mu}(\nu) < +\infty$ . Queremos mostrar que  $\nu \ll \pi$  e  $I_{\pi}(\nu) \leq \lambda_{\mu}(\nu)$ .

Note que se  $\eta \in \mathcal{M}_1(\mathbb{R}^d)$  e  $f$  é Borel mensurável e limitada então existem  $g_n$ ,  $n \geq 1$  em  $C_b(\mathbb{R}^d)$  tais que

$$\int_{\mathbb{R}^d} g_n d\eta \rightarrow \int f d\eta. \quad \text{Daí obtemos}$$

$$H(f) \stackrel{\text{def}}{=} \int f d\nu - \log \int e^f d\pi \leq \lambda_{\mu}(\nu) < +\infty \quad (1.30)$$

toda  $f \geq 0$  Borel -mensurável e limitada. Se  $\pi(A) = 0$  e tomamos  $f = n I_A$  temos

$$n\nu(A) \leq \lambda_{\mu}(\nu) \quad \text{para todo } n \text{ i.e., } \nu(A) = 0$$

de modo que  $\nu \ll \pi$ . Seja então  $g = \frac{d\nu}{d\pi}$  e basta ver que

$$\int g \log g d\pi = H(\log g) \leq \lambda_{\mu}(\nu) \quad (1.31)$$

Sejam  $g_n = g \mathbb{I}(1/n < g < n)$ . Então por (1.30) temos  $H(\log g_n) \leq \lambda_{\mu}(v)$ . Fazendo  $n \rightarrow \infty$  segue-se (1.31) e concluímos a prova de (i). ■

### Comentários e referências adicionais

Resultados como o Teorema 1 de Cramèr-Chernoff e Sanov têm muitas aplicações importantes em Estatística das quais já mencionamos uma definição natural de eficiência assintótica para certos testes de hipóteses. Neste sentido algumas referências são: Chernoff (1952), Bahadur (1971), Serfling (1980). Uma observação muito útil (Groeneboom et al. (1979)) consiste em obter, a partir do Teorema de Sanov, estimativas de grandes desvios para certas transformações da medida empírica  $T\left(\frac{1}{n} \sum \delta_{Y_i}\right)$  onde  $T: \mathbb{M}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^k$  é semicontínua inferiormente. Na verdade a hipótese sobre  $T$  pode ser enfraquecida permitindo  $T$  s.c.i. com relação a outra topologia que é mais forte do que a  $w^*$  (cf. Groeneboom et al. (1979)). A utilidade de tais observações consiste, por exemplo, no tratamento de estatísticas como "médias truncadas" e outras que se impõem pelo aspecto de robustez.

Bahadur-Zabell (1979) contém também vários exemplos interessantes onde as v.a.'s  $X_n$  tomam valores em espaços funcionais, i.e. são processos estocásticos. Exemplos importantes são processos gaussianos tais como o "movimento browniano", a folha browniana, o processo de Ornstein-Uhlenbeck, e outros. Na verda-

de os Teoremas 2.2 e 2.3 no próximo capítulo poderiam ser tratados desta forma, o que de certa forma unificaria as Teorias à Cramèr-Chernoff (que se referem a médias de variáveis i.i.d. e as estimativas básicas da Teoria de Freidlin e Wentzell sobre "pequenas perturbações aleatórias de sistemas determinísticos. Para este enfoque referimos a Azencott (1980), que trata de situações tecnicamente mais delicadas do que a que discutiremos no Capítulo 2.

Donsker e Varadhan (1975-76) tratam de grandes desvios para medidas empíricas  $\frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$  onde  $(Y_n)_{n \geq 1}$  é um processo de Markov. Também examinam processos a tempo contínuo i.e. trata-se de obter o comportamento assintótico das "medidas de ocupação"  $\mu_t$ , dadas por  $\mu_t(A) = \frac{1}{t} \int_0^t I_A(Y_s) ds$ , quando  $t \rightarrow +\infty$ . Extremamente importantes são as conexões com problemas de equações diferenciais parciais (Varadhan (1975)), mas, infelizmente, nosso pequeno curso não comporta uma discussão de toda esta teoria.

Uma consulta a Lanford (1961) é altamente recomendada; suas idéias básicas, lá apresentadas como uma mera "digressão" foram importantíssimas. Ademais é interessante notar como, mais uma vez, a Mecânica Estatística tem um papel importante no desenvolvimento da Teoria de processos estocásticos; isto aliás vem desde Boltzmann e Gibbs. Uma introdução ao estudo de sistemas Markovianos motivada por este enfoque é encontrada no texto "Int. aos Sist. Markovianos..." 5º SINAPE(1982) de A. Galves, A. Nogueira e M. Vares.

Finalmente uma observação mais técnica. A relação en-

tre  $\log \hat{\mu}(t)$  e  $\lambda_{\mu}(\cdot)$  é conhecida como dualidade de Legendre no contexto de funções convexas. Uma referência seria: R. Rockafellar. Convex Analysis. Princeton Univ. Press (1970).

Exercícios.

- 1.1 Seja  $\lambda(\cdot)$  definida por (1.2). Verifique que  $\lambda_{\mu}(\cdot)$  é convexa e semicontínua inferiormente, assim como a função  $\log \hat{\mu}: \mathbb{R} \rightarrow (-\infty, +\infty]$ .
- 1.2 Seja  $\mu$  a lei de uma v.a.  $X$ . Mostre que a transformada de Cramèr (1.2) da lei de  $Y = -X$  e  $Z = X-b$  é respectivamente  $\lambda_Y(x) = \lambda_{\mu}(-x)$  e  $\lambda_Z(x) = \lambda_{\mu}(x+b)$ . Conclua que (a) no Teorema 1.2 decorre de (1.7).
- 1.3 (a) Mostre que se  $X_1$  é v.a. Bernoulli com probabilidade de sucesso  $p$  ( $0 < p < 1$ ), então a transformada de Cramèr (dada por (1.2)) é:

$$\begin{aligned} \lambda(x) &= x \log(x/p) + (1-x) \log\left(\frac{1-x}{1-p}\right) \quad \text{se } 0 < x < 1 \\ &= -\log p \quad \text{se } x = 1 \\ &= -\log(1-p) \quad \text{se } x = 0 \\ &= +\infty \quad \text{se } x \notin [0,1] \end{aligned}$$

- (b) Calcule então  $\lambda_{\mu}(\cdot)$  quando  $\mu = p\delta_a + (1-p)\delta_b$  e  $a < b$  são reais, com  $0 < p < 1$ .



- (c) Ache  $\lambda_\mu$  quando  $\mu$  é a distribuição exponencial média  $\alpha^{-1}$ , i.e.,  $\mu(dx) = \alpha e^{-\alpha x} I_{(0,+\infty)}(x) dx$ .
- (d) Ache  $\lambda_\mu$  quando  $\mu$  é a distribuição normal com média  $\mu$  e variância  $\sigma^2$ .
- 1.4 Verifique a relação (1.11) e então deduza (1.10).
- 1.5 Mostre que se  $\mu(0,+\infty) > 0$  e  $\mu(-\infty,0) > 0$  então existem  $A, B$  positivos tais que  $\hat{\mu}(t) \geq A e^{B|t|}$  para todo  $t$ .
- 1.6 Seja  $X$  v.a. tal que  $E|X| < +\infty$ . (a) Se  $\mu$  é a lei de  $X$  mostre que  $\lambda_\mu(EX) = 0$ , (b) Motre que se  $[a,b]$  for o menor intervalo fechado  $I \subseteq [-\infty, +\infty]$  tal que  $P(X \in I) = 1$  então  $\lambda_\mu(\cdot)$  é finita e contínua em  $(a,b)$ , e  $\lambda_\mu(\bar{x}) = +\infty$   $x > b$  ou  $x < a$ . No caso  $a$  ou  $b$  finitos examine o comportamento nestes pontos. (Sugestão: utilize o Teorema 1.2).
- 1.7 Seja  $\mu$  probabilidade em  $\mathbb{R}$  com  $\int (x) \mu(dx) < +\infty$ . Não é difícil provar que  $\hat{\mu}(t) = +\infty$  para todo  $t > 0$  equivale a  $\lambda_\mu(x) = 0$  para todo  $x \geq 0$ . Utilize isto para dar exemplo em que  $\lambda_\mu \equiv 0$ .
- 1.8 Suponha  $X$  v.a. com lei  $\mu$  tal que  $\hat{\mu}(t) < +\infty$  para todo  $t \in (-\delta, \delta)$ , para algum  $\delta > 0$ . Mostre que  $\lambda''(EX) > 0$ . Que significa isto para as estimativas obtidas?

1.9 Utilize o resultado citado no final da seção 1.2 conclua que se  $\mu$  for uma probabilidade em  $\mathbb{R}^k$  com  $\hat{\mu}(t) < +\infty$  para todo  $t$  em alguma vizinhança da origem, então para todo  $A$  Boreliano temos

$$-\Lambda(\overset{\circ}{A}) \leq \underline{\lim}_{n \rightarrow +\infty} \frac{1}{n} \log P(X_n \in A) \leq \overline{\lim}_{n \rightarrow +\infty} \frac{1}{n} \log P(X_n \in A) \leq \Lambda(\bar{A})$$

onde  $X_1, X_2, \dots$  são i.i.d. com lei  $\mu$  e  $\Lambda$  é definido por (1.19).

1.10 Dê exemplo nas condições do exercício 8, de um boreliano  $A$  tal que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\bar{X}_n \in A) = \iota(A)$$

exista com  $-\Lambda(A) = \iota(A)$ , mas  $\Lambda(\overset{\circ}{A}) \neq \Lambda(\bar{A})$ .

(Sugestão: utilize exemplo do exercício 3.)

CAPÍTULO 2

PEQUENAS PERTURBAÇÕES ALEATÓRIAS DE SISTEMAS DINÂMICOS

— TEORIA DE FREIDLIN E WENTZELL —

§2.1 Introdução do modelo. Estimativas básicas

Neste capítulo estaremos considerando processos de difusão em  $\mathbb{R}^d$ ,  $X_\epsilon^x(\cdot)$ , definidos por equação diferencial estocástica

$$\begin{aligned}dX_\epsilon^x(t) &= b(X_\epsilon^x(t))dt + \epsilon dW_t \\ X_\epsilon^x(0) &= x\end{aligned}\tag{2.1}$$

onde  $\epsilon > 0$ ,  $x \in \mathbb{R}^d$  é o ponto inicial,  $(W_t)$  é um movimento Browniano (padrão) em  $\mathbb{R}^d$  i.e.  $W_t = (W_t^1, \dots, W_t^d)$   $t \geq 0$  onde os  $W^i$  são movimentos brownianos reais independentes (padrão) e o campo de vetores  $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$  verifica:

$$\begin{aligned}(a) \quad &|b(x) - b(y)| \leq K|x-y|; \quad x, y \in \mathbb{R}^d \\ (b) \quad &|b(x)| \leq K(1+|x|); \quad x \in \mathbb{R}^d\end{aligned}\tag{2.2}$$

sendo  $|\cdot|$  a norma Euclídeana em  $\mathbb{R}^d$  e  $K$  uma constante positiva e finita.

Como é bem sabido, as trajetórias  $t \rightarrow W_t(\omega)$  não são diferenciáveis (com probabilidade um) e (2.1) é entendida como

uma equação integral

$$X_{\epsilon}^x(t) = x + \int_0^t b(X_{\epsilon}^x(s)) ds + \epsilon W_t, \quad t \geq 0.$$

[Para as propriedades básicas do movimento Browniano há muitas referências entre os livros-textos de processos estocásticos existentes (e.g. Breiman (1968)). Para as equações diferenciais estocásticas de Ito citamos, por exemplo, o Capítulo 1 do livro de Ikeda e Watanabe embora hajam muitas outras referências disponíveis.]

Estamos interessados na análise destes processos quando  $\epsilon$  é pequeno, e então  $X_{\epsilon}^x(\cdot)$  pode ser pensado como uma "pequena" perturbação estocástica do sistema dinâmico  $X^x(\cdot)$ , solução ( $C^1$ ) de

$$\begin{aligned} \dot{x}(t) &= b(x(t)) \\ x(0) &= x. \end{aligned} \tag{2.3}$$

De fato, se observarmos  $X_{\epsilon}^x(\cdot)$  durante um intervalo limitado  $[0, T]$ , ( $T < +\infty$ ), com grande probabilidade veremos uma trajetória que embora oscilando muito mantém-se próxima à de  $X^x(\cdot)$ . Mais precisamente, é fácil constatar que dado  $\delta > 0$

$$P \left( \sup_{t \leq T} |X_{\epsilon}^x(t) - X^x(t)| > \delta \right) \xrightarrow{\epsilon \downarrow 0} 0$$

para todo  $x \in \mathbb{R}^d$ . Na verdade, a convergência é uniforme em  $x$  e ocorre rapidamente (decaimento exponencial). Este fato, muito simples, e que justifica o adjetivo "pequeno" acima, decorre das

seguintes observações: por (2.1) e (2.3) temos:

$$X_{\epsilon}^x(t) - X^x(t) = \int_0^t [b(X_{\epsilon}^x(s)) - b(X^x(s))] ds + \epsilon W_t$$

de modo que por (2.2)

$$|X_{\epsilon}^x(t) - X^x(t)| \leq K \int_0^t |X_{\epsilon}^x(s) - X^x(s)| ds + \epsilon |W_t|$$

e então, pela desigualdade de Gronwall:

$$\sup_{t \leq T} |X_{\epsilon}^x(t) - X^x(t)| \leq \epsilon e^{KT} \sup_{t \leq T} |W_t|. \quad (2.5)$$

Por outro lado, como é bem conhecido (exercício 2.1):

$$\begin{aligned} P(\sup_{t \leq T} |W_t| > \eta) &\leq dP(\sup_{t \leq T} |W_t^1| > \eta/d) \leq \\ &\leq 2dP(\sup_{t \leq T} W_t^1 > \eta/d) \\ &\leq 4dP(W_T^1 > \eta/d) \\ &\leq 4dP(W_1^1 > \eta/d\sqrt{T}), \end{aligned} \quad (2.6)$$

onde  $W_1^1$  tem lei normal com média zero e variância 1, i.e.,

$\mu(du) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ . Por (2.5), (2.6) e como

$$\int_x^{+\infty} e^{-u^2/2} du \leq \frac{1}{x} e^{-x^2/2} \quad \text{se } x > 0 \quad (2.7)$$

obtemos

$$P(\sup_{t \leq T} |X_{\epsilon}^x(t) - X^x(t)| > \delta) \leq C_1 e^{-C/\epsilon^2} \quad (2.8)$$

onde  $C, C_1$  são constantes positivas (dependem de  $T, K, \delta$  e  $d$ ).

Se  $C([0, T], \mathbb{R}^d) \stackrel{\text{def}}{=} \{y: [0, T] \rightarrow \mathbb{R}^d / y \text{ é contínua}\}$  com a métrica  $\rho_T(x, y) = \sup_{0 \leq t \leq T} |x(t) - y(t)|$ , (2.8) fica:

$$P(\rho_T(X_\epsilon^x, X^x) > \delta) \leq C_1 e^{-C/\epsilon^2} \quad (2.8)'$$

Ou seja, se  $A$  for um Boreliano de  $C[0, T] \stackrel{\text{not}}{=} C([0, T], \mathbb{R}^d)$ , a uma distância positiva da trajetória determinística  $X^x(\cdot)$ , então a probabilidade de  $X_\epsilon^x(\cdot)$  pertencer a  $A$  decai exponencialmente a zero quando  $\epsilon \downarrow 0$ , ou seja, o evento  $[X_\epsilon^x \in A]$  representa uma situação de grande desvio ( $\epsilon \downarrow 0$ ). A pergunta que se coloca é de como melhor avaliar as probabilidades de tais eventos raros de modo que possamos comparar  $P[X_\epsilon^x \in A]$  e  $P[X_\epsilon^x \in B]$ , para  $A$  e  $B$  ambos afastados da trajetória determinística  $(X^x(t): 0 \leq t \leq T)$ .

A preocupação com um tal estudo pode até parecer estranha à primeira vista, já que de modo geral estamos interessados no que ocorre com probabilidade um, ou tendendo a um. Entretanto os problemas mais interessantes com relação a esta classe de processos referem-se ao comportamento ao longo de intervalos de tempo que podem (e devem) se tornar arbitrariamente grandes quando  $\epsilon \downarrow 0$ . Exemplos típicos são as questões referentes a medidas invariantes, ou o problema de como  $X_\epsilon(\cdot)$  escapa de um domínio  $D$  invariante para  $X(\cdot)$ , quando a trajetória determinística seria atraída a um ponto, ou a um compacto em  $D$ . Neste caso, o tempo

necessário até esta saída é outra variável importante. O que acontece então é que no estudo do comportamento típico de  $X_\epsilon(\cdot)$  em intervalos cujo comprimento tende a infinito, será indispensável podermos avaliar e comparar as probabilidades de eventos raros; aí "grandes flutuações" com relação ao comportamento mais provável em intervalos  $[0, T]$  com  $T$  fixado, realmente ocorrerão, originando um comportamento totalmente distinto do sistema determinístico.

O método que passamos a descrever foi desenvolvido por Freidlin e Wentzell (1970, 1972, 1984)<sup>(1)</sup>, gozando de uma profunda analogia com algo bem mais antigo; o método de Laplace para estimar integrais da forma

$$\int_a^b e^{-\epsilon^{-1}f(x)} g(x) dx \quad (2.9)$$

quando  $\epsilon \downarrow 0$ , onde  $f, g$  são funções reais contínuas e  $g > 0$ . De certa forma sua teoria requer uma "versão infinito - dimensional" do método de Laplace. No caso de  $f$  assumir o mínimo em um único ponto  $x_0$ , e sendo  $U \subseteq [a, b]$  vizinhança de  $x_0$ , é muito simples verificar que, quando  $\epsilon \downarrow 0$

$$\int_{[a, b] \setminus U} e^{-\epsilon^{-1}f(x)} g(x) dx / \int_U e^{-\epsilon^{-1}f(x)} g(x) dx \xrightarrow{\text{(exponencialmente)}} 0 \quad (2.10)$$

De fato, sejam  $m, M$  t.q.  $0 < m \leq g(x) \leq M$  em  $[a, b]$  e

---

(1) As referências aos trabalhos de Freidlin e Wentzell serão indicadas por F-W(...).

seja  $\eta > 0$  tal que  $\min_{[a,b] \setminus U} f(x) \geq f(x_0) + \eta$ ; temos então

$$\int_{[a,b] \setminus U} e^{-\epsilon^{-1}f(x)} g(x) dx \leq (b-a)M e^{-\frac{f(x_0)+\eta}{\epsilon}}; \quad (2.11)$$

por outro lado, se  $(x_0 - \delta, x_0 + \delta) \subset U$  e  $f(y) \leq f(x_0) + \eta/2$  para  $|y - x_0| \leq \delta$ , teremos

$$\int_U e^{-\epsilon^{-1}f(x)} g(x) dx \geq 2\delta m e^{-\frac{f(x_0) + \eta/2}{\epsilon}} \quad (2.12)$$

de modo que a razão em (2.10) será majorada por  $\frac{(b-a)M}{2\delta m} e^{-\eta/2\epsilon}$ , verificando (2.10).

Obs. Na verdade o que temos acima é

$$\int_a^b e^{-\epsilon^{-1}f(x)} g(x) dx \sim \epsilon^{-1}f(x_0) \quad \text{quando } \epsilon \rightarrow 0.$$

A existência de uma coisa análoga para o cálculo de  $P(X_\epsilon^x \in A)$ , onde  $A$  é um Boreliano de  $C[0, T]$  a uma distância positiva da trajetória  $X^x(\cdot)$ , permitiria investigar quais porções de  $A$  mais contribuem para esta probabilidade. i.e., permitiria saber da existência de função  $\varphi^* \in A$  tal que se  $U \subset A$  é vizinhança de  $\varphi^*$  então

$$P(X_\epsilon^x \in U \mid X_\epsilon^x \in A) \rightarrow 1 \quad \text{quando } \epsilon \rightarrow 0.$$

Para tanto, a idéia básica de Freidlin e Wentzell con-



siste em determinar a existência de funcional  $S$  em  $C[0, T]$  tal que  $P(\mu_T(X_\epsilon, \varphi) \leq \delta)$  se comporte como  $\exp(-\epsilon^{-2} S(\varphi))$ ; isto sugeria então uma maior concentração da probabilidade em vizinhança das funções que minimizam  $S(\cdot)$ . Note que por analogia com o método de Laplace precisaríamos:

- (a) cota inferior para  $P(X_\epsilon \in B(\varphi))$
- (b) cota superior para  $P(X_\epsilon \in A \setminus B(\varphi))$

de modo análogo a (2.11) e (2.12), onde  $B(\varphi)$  denota vizinhança de  $\varphi$  contida em  $A$ . Nos Teoremas 2.1 e 2.2 abaixo, verificaremos que isto funciona em um caso bem particular: quando  $b=0$ , i.e.,  $X_\epsilon^x = x + \epsilon W_t$  sendo  $S(\varphi)$  dado por

$$S(\varphi) = \frac{1}{2} \int_0^T |\dot{\varphi}(s)|^2 ds \quad \text{se } \varphi \text{ absolutamente contínua}$$

e  $\varphi(0) = x$

=  $+\infty$  caso contrário.

O caso geral quando  $b(\cdot)$  verifica (2.2) será obtido facilmente, já que neste caso a transformação que leva  $(\epsilon W_t: t \leq T)$  em  $(X_\epsilon(t): t \leq T)$  é contínua. É esta, na verdade, a razão para nos limitarmos, neste curso, à classe de difusões onde  $B(\cdot)$  verifica (2.2), não depende de  $\epsilon$ , e o coeficiente de difusão é constante. Resultados mais gerais, envolvendo difusões do tipo

$$dY_\epsilon^x(t) = b_\epsilon(Y_\epsilon^x(t))dt + \epsilon \sigma(Y_\epsilon^x(t))dW_t,$$

$$Y_\epsilon^x = x$$

podem também ser obtidos (F-W (1970), Azencott (1980)), mas para os pontos que desejamos enfatizar no curso não há necessidade de tal sofisticação; as idéias básicas já estão contidas na situação que estudaremos.

No trabalho de Feynman em Mecânica Quântica (Feynman, Hibbs (1965)<sup>(\*)</sup>), existem construções em certo sentido semelhantes. Lá aparece então um funcional, chamado funcional de ação, e que na descrição do sistema quântico tem um papel de certo modo análogo ao  $S(\varphi)$  descrito acima na Teoria de Freidlin e Wentzell. Por esta razão, eles chamaram  $e^{-2} S(\varphi)$  de funcional de ação para  $X_\epsilon(\cdot)$ .

Antes de provarmos os resultados sobre  $X_\epsilon(t) = e^{W_t}$  talvez seja conveniente definir o que foi dito acima em um contexto mais geral. Isto será de utilidade, pois evitará trabalho adicional para tratar das difusões  $X_\epsilon^x(\cdot)$  a que nos propusemos, fornecendo também um grande número de aplicações bem como permitindo "vislumbrarmos" o potencial da teoria.

Antes de tudo, nossas afirmações referem-se a probabilidades de eventos  $[X_\epsilon \in A]$  onde  $A$  é um conjunto de trajetórias (Boreliano de  $C[0, T]$ ). Ou seja, referem-se às medidas de probabilidade  $P_\epsilon^x$  em  $C[0, T]$ , onde  $P_\epsilon^x$  é a lei do processo  $X_\epsilon^x(\cdot)$ . Deste modo o problema pode ser colocado no cenário de uma família de medidas de probabilidade  $(\mu_\epsilon; \epsilon > 0)$  em algum espaço métrico  $(M, \rho)$ . A idéia de  $\mu_\epsilon$  representar uma pequena perturbação estocástica de um "sistema determinístico" ( $M$  espaço funcional) é expressa pela existência de  $x_0 \in M$  tal que

(\*) Quantum Mechanics and Path Integrals, McGraw Hill.

$\mu_\epsilon \{x: \rho(x, x_0) > \delta\} \rightarrow 0$  para cada  $\delta > 0$ .

Definição 2.1. Nas condições acima, se  $\lambda_\epsilon \geq 0$  e  $\lambda_\epsilon \nearrow +\infty$  quando  $\epsilon \downarrow 0$  e  $S: M \rightarrow [0, +\infty]$ , diremos que  $\lambda_\epsilon S(\cdot)$  é funcional de ação para  $\{\mu_\epsilon\}$ , se:

(O)  $\forall s > 0 \quad \bar{\Phi}(s) = \{p: S(p) \leq s\}$  é compacto;

(I)  $\forall \delta > 0, \forall \gamma > 0, \forall p \in M$ , existe  $\epsilon_0 > 0$  t.q.

$$\mu_\epsilon \{q: \rho(p, q) < \delta\} \geq \exp[-\lambda(\epsilon)(S(p) + \gamma)];$$

(II)  $\forall \delta > 0, \forall \gamma > 0, \forall s > 0, \exists \epsilon_0 > 0$  t.q.

$$\mu_\epsilon \{q: \rho(q, \bar{\Phi}(s)) \geq \delta\} \leq \exp(-\lambda_\epsilon(s-\gamma)) \text{ se } \epsilon \leq \epsilon_0.$$

Tipicamente  $M$  será um espaço funcional ( $C[0, T]$ , no nosso caso) e  $(Y_\epsilon)_\epsilon$  uma família de processos com trajetórias em  $M$ , sendo então  $\mu_\epsilon$  a lei (ou distribuição) de  $Y_\epsilon$ .

É claro que  $\lambda_\epsilon$  e  $S(\cdot)$  não estão univocamente determinados. Entretanto pode-se provar que uma vez fixado  $\lambda_\epsilon$  então o funcional de ação  $S(\cdot)$  fica univocamente determinado (Exercício 2.2).

A condição (O) não é uma technicalidade à toa<sup>(1)</sup>; garante que se  $A$  for um conjunto fechado e disjunto de  $\bar{\Phi}(s)$  (i.e. com  $S(p) > s$  para cada  $p \in A$ ), então existe  $\delta > 0$  tal que  $\mu_\epsilon(A, \bar{\Phi}(s)) \geq \delta$ , de modo que (II) fornece cota supe-

(1) Compare com comentários no final de § 1.2.

rior para  $\mu_\epsilon(A)$ . O leitor logo encontrará outra maneira de ver isto pois (0) garante que  $S$  atinge o mínimo sobre qualquer conjunto fechado (Exercício 2.3). Observemos também que no caso  $X$  completo, (0) equivale aos seguintes requisitos: para cada  $\rho > 0$

(0-a)  $\tilde{\varphi}(s)$  é relativamente compacto

(0-b)  $\tilde{\varphi}(s)$  é fechado (ou, equivalentemente,  $S(\cdot)$  é semicontínua inferiormente).

(ver Exercício 2.4).

Voltemos a considerar os processos  $X_\epsilon^x(\cdot)$ . Por enquanto examinaremos o caso  $X_\epsilon(t) = \epsilon W_t$  i.e.  $b \equiv 0$  (aqui  $x$  entra apenas para uma translação e por isto faremos  $x = 0$ ).

Definimos, para cada  $T > 0$  o funcional  $S_T(\cdot)$  em  $C[0, T]$ , via:

$$S_T(\varphi) = \frac{1}{2} \int_0^T |\dot{\varphi}(s)|^2 ds, \quad \text{se } \varphi(\cdot) \text{ for absolutamente} \\ \text{contínua e } \varphi(0) = 0 \quad (2.13)$$

=  $+\infty$ , caso contrário.

Provaremos então que  $\epsilon^{-2} S_T$  é funcional de ação para  $(\epsilon W_t: 0 \leq t \leq T)$  conforme o seguinte:

Teorema 2.2. (F-W (1970), (1984))

(a) Para todo  $\delta > 0$ ,  $\gamma > 0$ , e  $0 < K_0 < +\infty$ , existe  $\epsilon_0 > 0$

tal que  $P(\rho_T(X_\epsilon, \varphi) < \delta) \geq \exp \left\{ -\frac{1}{\epsilon^2} (S_T(\varphi) + \gamma) \right\}$  se  
 $\epsilon < \epsilon_0$ ,  $\varphi \in C[0, T]$  e  $T + S_T(\varphi) \leq K_0$

(b) Se  $\mathfrak{F}_T(s) = \{\varphi \in C[0, T]: \varphi(0) = 0 \text{ e } S_T(\varphi) \leq s\}$ , então para  
todo  $\delta > 0$ ,  $\gamma > 0$  e  $s_0 > 0$  existe  $\epsilon_0 > 0$

$P(\rho_T(X^\epsilon, \mathfrak{F}(s)) \geq \delta) \leq \exp \left\{ -\left(\frac{s-\gamma}{\epsilon^2}\right) \right\}$  se  $\epsilon < \epsilon_0$ , e  $s \leq s_0$ .

(c)  $\mathfrak{F}_T(s)$  definido em (b) é compacto em  $C[0, T]$ , para cada  
 $s < +\infty$ .

### Observações:

1. Em muitos casos  $T$  estará fixo e não havendo possibilidade de confusão será eliminado da notação.
2. A condição contida no item (b) do Teorema 2.2 é equivalente à condição (II) da Definição 2.1 (embora possa parecer mais forte à primeira vista), conforme o leitor verificará facilmente.

### Prova:

(a) Como  $S_T(\varphi) < +\infty$  temos que  $\varphi$  é absolutamente contínua,  
 $\varphi(0) = 0$ , i.e.,  $\varphi(t) = \int_0^t \dot{\varphi}(s) ds$ , e  $\int_0^T |\dot{\varphi}(s)|^2 ds =$   
 $= 2S_T(\varphi) < +\infty$ .

Dada uma tal função  $\varphi$  denotemos por  $Z_\epsilon(t)$ ,  $0 \leq t \leq T$  o processo  $W_t - \epsilon^{-1} \varphi_t$ , de modo que  $\left[ (P_T(X^\epsilon, \varphi) < \delta) \right] = \left[ \sup_{t \leq T} \epsilon |Z_\epsilon(t)| < \delta \right]$ . O Teorema de Cameron-Martin-Girsanov nos diz que a lei de  $(W_t - \epsilon^{-1} \varphi_t: 0 \leq t \leq T)$  é absolutamente contínua com relação à de  $(W_t: 0 \leq t \leq T)$  e a derivada de Radon-Nikodym é

$$L(w) = \exp\left(-\epsilon^{-1} \int_0^T \langle \varphi(s), dW_s \rangle - \frac{\epsilon^{-2}}{2} \int_0^T |\dot{\varphi}(s)|^2 ds\right).$$

Dai obtemos facilmente

$$\begin{aligned} P(P_T(X^\epsilon, \varphi) < \delta) &= P\left(\sup_{t \leq T} \epsilon |Z_\epsilon(t)| < \delta\right) = \\ &= \int_{\left[\sup_{t \leq T} |\epsilon W_t| < \delta\right]} L(w) P(dw) = \quad (2.14) \\ &= \exp(-\epsilon^{-2} S(\varphi)) \int_{\left[\sup_{t \leq T} |\epsilon W_t| < \delta\right]} \exp\left\{-\epsilon^{-1} \int_0^T \langle \varphi(s), dW_s \rangle\right\} P(dw) \end{aligned}$$

Dados  $\delta > 0$  e  $K_0 > 0$

$$P\left(\sup_{t \leq K_0} |\epsilon W_t| \geq \delta\right) \longrightarrow 0 \quad \text{quando} \quad \epsilon \longrightarrow 0$$

e portanto tomemos  $\epsilon'_0$  tal que

$$P\left(\sup_{t \leq K_0} |\epsilon W_t| < \delta\right) \geq \frac{2}{3} \quad \text{se} \quad \epsilon \leq \epsilon'_0.$$

Por outro lado, se  $c > 0$

$$\begin{aligned}
 P(\epsilon^{-1} \int_0^T \langle \varphi(s), dW_s \rangle \geq c \epsilon^{-1} \sqrt{S_T(\varphi)}) \\
 \leq P\left(\left(\int_0^T \langle \varphi(s), dW_s \rangle\right)^2 > c^2 S_T(\varphi)\right) \\
 \leq 2/c^2 \quad \text{pela desigualdade de Markov,}
 \end{aligned}$$

pois  $E\left(\int_0^T \langle \varphi(s), dW_s \rangle\right)^2 = 2S_T(\varphi)$ . Tomando por exemplo  $c = \sqrt{6}$

obtemos

$$1/3 \leq P\left[\sup_{t \leq T} |\epsilon W_t| < \delta, \quad \epsilon^{-1} \int_0^T \langle \varphi(s), dW_s \rangle < c \epsilon^{-1} \sqrt{S_T(\varphi)}\right]$$

e portanto por (2.14) temos:

$$P(\rho_T(X_\epsilon, \varphi) < \delta) \geq \frac{1}{3} \exp(-\epsilon^{-2} \cdot S_T(\varphi) - c \epsilon^{-1} \sqrt{S_T(\varphi)})$$

se  $\epsilon \leq \epsilon'_0$  e  $T \leq K_0$ . Agora, dado  $\gamma > 0$ , tomemos  $\epsilon_0 \leq \epsilon'_0$  tal que

$$\frac{1}{3} \exp(-\epsilon^{-1} \sqrt{6K_0}) > \exp(-\gamma \epsilon^{-2}) \quad \text{se } \epsilon \leq \epsilon_0.$$

Com isto obtemos a prova de (a).

(b) Seja  $N$  inteiro  $\geq 1$  e consideremos para cada  $\epsilon > 0$  o processo estocástico  $p_\epsilon^N$ , cujas trajetórias consistem em linhas poligonais ligando os pontos  $(\frac{kT}{N}, \epsilon W_{kT/N})$ ,  $k = 0, 1, \dots, N$ . Então, para cada  $s, \delta, \gamma > 0$  e para cada  $N \geq 1$ :

(2.15)

$$P(\rho_T(X_\epsilon, \bar{\phi}(s)) \geq \delta) \leq P(\rho_T(X_\epsilon, p_\epsilon^N) > \delta) + P(S_T(p_\epsilon^N) > s).$$

Estimemos pois cada um dos termos no lado direito de (2.15). Temos:

$$\begin{aligned} P(\rho_T(X_\epsilon, p_\epsilon^N) > \delta) &\leq NP \left( \sup_{t \leq T/N} |\epsilon W_t - \epsilon \frac{tN}{T} W_{T/N}| \geq \delta \right) \\ &\leq NP \left( \sup_{t \leq T/N} |W_t| > \delta/2\epsilon \right) \\ &\leq 4dN P(W_1^1 > \delta\sqrt{N}/2d\epsilon\sqrt{T}), \end{aligned}$$

onde na primeira desigualdade usamos a estacionariedade dos incrementos de  $(W_t)$  o que implica serem as variáveis  $(i = 1, \dots, N)$   $\sup_{(i-1)T/N \leq t \leq iT/N} |\epsilon W_t - p_\epsilon^N(t)|$  identicamente distribuídas; a segunda desigualdade é óbvia e a terceira vem de (2.6). Utilizando-se agora a desigualdade (2.7) obtemos (como em (2.8))

$$P(\rho_T(X_\epsilon, p_\epsilon^N) > \delta) \leq \frac{4dN}{\sqrt{2\pi}} \frac{2\epsilon d}{\delta} \sqrt{\frac{T}{N}} \cdot e^{-\frac{\delta^2 N}{8d^2 T \epsilon^2}} \quad (2.16)$$

e daí

$$\begin{aligned} P(\rho_T(X_\epsilon, p_\epsilon^N) > \delta) &\leq \frac{8d^2}{\delta} \sqrt{\frac{NT}{\pi}} \epsilon e^{-\frac{1}{8} \frac{\delta^2 N}{d^2 T \epsilon^2}} \\ &\leq \frac{8d^2}{\delta} \sqrt{\frac{NT}{\pi}} \epsilon e^{-s_0/2\epsilon^2} \leq \frac{1}{2} e^{-\frac{1}{2} \frac{(s-\gamma)}{\epsilon^2}} \end{aligned} \quad (2.17a)$$

se  $N > 4d^2 T s_0 / \delta^2$ , e  $\epsilon_0 > 0$  for suficientemente pequeno,



para todo  $s \leq s_0$ , se  $\epsilon \leq \epsilon_0$ . Por outro lado

$$S_T(p_\epsilon^N) = \frac{\epsilon}{2} \sum_{k=1}^N N |W_{KT/N} - W_{(k-1)T/N}|^2,$$

sendo que  $\sqrt{N}(W_{kT/N} - W_{(k-1)T/N})$  são vetores aleatórios independentes e identicamente distribuídos; a lei de cada um é Normal d-dimensional, o vetor de médias é 0 e a matriz de covariâncias é a identidade. Logo

$$S_T(p_\epsilon^N) \text{ tem a mesma distribuição que } \frac{\epsilon}{2} \sum_{i=1}^{dN} |Z_i|^2$$

onde  $Z_1, \dots, Z_{dN}$  são v.a.'s reais i.i.d. com distribuição Normal de média 0 e variância 1. Ademais, se  $\theta < 1$

$$E e^{\frac{\theta Z_1^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\frac{\theta-1}{2} x^2} dx = C_\theta < +\infty$$

(Exercício 2.5) e então

$$\begin{aligned} P(S_T(p_\epsilon^N) > s) &= P\left(e^{\frac{\theta}{2} \sum_{i=1}^{dN} |Z_i|^2} > e^{\frac{\theta s}{\epsilon^2}}\right) \\ &\leq e^{-\frac{\theta s}{\epsilon^2}} (C_\theta)^{dN} \\ &\leq e^{-\frac{s}{\epsilon^2}} e^{\frac{(1-\theta)s}{\epsilon^2}} C_\theta^{dN}. \end{aligned}$$

Mas então fixando  $N$  conveniente pela desigualdade anterior e tomando  $\theta$  tal que  $1-\theta < \gamma/2s_0$  teremos  $P(S_T(p_\epsilon^N) > s) \leq e^{-\frac{s}{\epsilon^2}} e^{\gamma/\epsilon^2}$ . Tomando  $\epsilon'_0 > 0$  t.q.  $C_\theta^{dN} e^{\gamma/\epsilon^2} \leq \frac{1}{2} e^{\gamma/\epsilon^2}$  para  $\epsilon \leq \epsilon'_0$  teremos:

$$P(S_T(P_\epsilon^N) > s) \leq \frac{1}{2} e^{-\frac{(s-\gamma)}{\epsilon^2}} \quad (2.17b)$$

para todo  $s \leq s_0$  e  $\epsilon \leq \epsilon'_0$ . Então, (b) segue de (2.15), (2.17a) e (2.17b).

O ítem (c) decorre facilmente do Teorema de Arzelà-Ascoli e da semicontinuidade inferior de  $S_T(\cdot)$ . (Exercício 2.6).

O caso de processos de difusão  $X_\epsilon^x$  dados pela equação (2.1), onde o campo de vetores satisfaz (2.2), pode ser tratado semelhantemente ao que foi feito acima mas, talvez seja mais prático notar duas coisas: (a) a solução  $(X_\epsilon^x(\cdot))$  de (2.1) pode ser obtida por uma transformação contínua das trajetórias de  $(\epsilon W_t)$ ; (b) se temos  $\varphi$  uma transformação contínua entre dois espaços métricos  $M, \tilde{M}$  e em  $M$  temos uma família de probabilidades  $(\mu_\epsilon)$  descrita por um funcional de ação  $\lambda_\epsilon S$  então as medidas induzidas em  $\tilde{M}$  pela  $\varphi$  são descritas por um funcional de ação obtido do anterior por uma regra simples.

O Teorema 2.3 abaixo verifica (b).

Teorema 2.3. Seja  $(\mu_\epsilon)$  uma família de probabilidades em um espaço métrico  $(M, \rho)$  e supomos que o comportamento quando  $\epsilon \downarrow 0$  é descrito (no sentido da Definição 2.1) pelo funcional de ação  $\lambda_\epsilon S(\cdot)$ . Se  $B$  é uma transformação contínua de  $(M, \rho)$  em outro espaço métrico  $(\tilde{M}, \tilde{\rho})$  e  $\nu_\epsilon = \mu_\epsilon \circ B^{-1}$  então  $\lambda_\epsilon \tilde{S}(\cdot)$  define um funcional de ação para  $\nu_\epsilon$  (quando  $\epsilon \downarrow 0$ ), onde

$$\begin{aligned}\tilde{S}(y) &= \inf\{S(x) : B(x) = y\} \\ &= +\infty \text{ se } y \notin B(M).\end{aligned}$$

Prova: Fica como exercício, já que a verificação das condições (0), (I) e (II) é simples rotina. (Exercício 2.7).

Proposição 2.4. Suponhamos que  $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$  verifique (2.2). Podemos então definir, para cada  $x \in \mathbb{R}^d$ , a transformação  $B_x: C[0, T] \rightarrow C[0, T]$  via  $B_x \varphi = \psi$  se e só se  $\psi$  é a (única) solução de

$$\psi(t) = x + \int_0^t b(\psi(s)) ds + \varphi(t), \quad 0 \leq t \leq T \quad (2.18)$$

Ademais  $B_x$  é injetiva e

$$B_x^{-1} \psi(t) = \psi(t) - \int_0^t b(\psi(s)) ds - x \quad (2.19)$$

e, se  $K$  é a constante de Lipschitz para  $b(\cdot)$  então  $e^{KT}$  é constante de Lipschitz para  $B_x$ , independentemente de  $x \in \mathbb{R}^d$ , i.e.,

$$\rho_T(B_x \varphi, B_x \tilde{\varphi}) \leq e^{KT} \rho_T(\varphi, \tilde{\varphi}) \quad (2.20)$$

Prova: (2.2) garante a existência e unicidade da solução de (2.18). (2.19) é imediata da definição de  $B_x$  e (2.20) decorre da desigualdade de Gronwall aplicada a

$$|B_x \varphi(t) - B_x \tilde{\varphi}(t)| \leq K \int_0^t |B_x \varphi(s) - B_x \tilde{\varphi}(s)| ds + |\varphi(t) - \tilde{\varphi}(t)|$$

para todo  $t \in [0, T]$ , o que por sua vez, vem de (2.2) e (2.18). ■

Como consequência desta proposição e dos Teoremas 2.2 e 2.3 podemos definir um funcional de ação para os processos de difusão  $X_\epsilon^x(\cdot)$ , obtidos como solução da equação diferencial estocástica (2.1). De fato se  $(W_t)_{t \leq T}$  é um Movimento Browniano padrão em certo espaço de probabilidade  $(\Omega, \mathcal{G}, P)$  então

$$X_\epsilon^x(\omega, \cdot) = B_x(\epsilon W_\cdot(\omega))$$

é o processo de difusão que nos interessa, i.e. se  $\tilde{P}_\epsilon$  é a lei de  $(\epsilon W_t: 0 \leq t \leq T)$  em  $C[0, T]$  então  $P_\epsilon^x = \tilde{P}_\epsilon \circ B_x^{-1}$  é a lei do processo em questão. Mas então é imediato dos Teoremas 2.2 e 2.3 e Proposição 2.4, que definindo:

$$S_T^x(\varphi) = \frac{1}{2} \int_0^T |\dot{\varphi}(s) - b(\varphi(s))|^2 ds \quad \text{se } \varphi(0) = x \text{ e } \varphi \text{ absolutamente contínua}$$

=  $+\infty$  caso contrário,

então  $\epsilon^{-2} S_T^x(\cdot)$  é funcional de ação para  $(P_\epsilon^x)_{\epsilon > 0}$ , no sentido da Definição 2.1.

Uma pergunta bem natural refere-se à possibilidade das estimativas dadas por (I) e (II) independermos de  $x$ . Nos casos que estamos examinando realmente há uniformidade em  $x$ . Para melhor precisarmos esta afirmação comecemos por eliminar o ponto inicial  $x$  da nossa definição do funcional de ação, colocando sim-

plesmente

$$S_T(\varphi) = \frac{1}{2} \int_0^T |\dot{\varphi}(s) - b(\varphi(s))|^2 ds \quad \text{se } \varphi \text{ absolutamente cont nua} \quad (2.21)$$

= +∞ caso contr rio ,

definindo ent o:

$$\Phi_T(s, x) = \{\varphi \in C[0, T]: \varphi(0) = x \text{ e } S_T(\varphi) \leq s\}.$$

(Omitiremos o T quando n o causar confus o.)

Como a constante de Lipschitz em  $B_x$  pode ser tomada independentemente de x,   f cil concluir que (I) e (II) devem valer de modo uniforme em x. (O) tamb m pode ser refor ada. Mais precisamente, vale o seguinte:

Teorema 2.5. Suponhamos que  $b(\cdot)$  verifique (2.2) e seja  $X_\epsilon^x(\cdot)$  a solu o da equa o diferencial estoc stica (2.1). Se para cada  $T > 0$ ,  $S_T$    definida via (2.21), temos:

(O)<sub>u</sub> Para cada K compacto  $\subseteq \mathbb{R}^d$ , o conjunto  $\{\varphi: \varphi(0) \in K \text{ e } S_T(\varphi) \leq s\}$    compacto em  $C[0, T]$ .

(I)<sub>u</sub> Para cada  $\delta > 0$ ,  $\gamma > 0$  e  $s_0 > 0$  existe  $\epsilon_0 > 0$  tal que

$$P(\rho_T(X_\epsilon^x(\cdot), \varphi) < \delta) \geq \exp(-\epsilon^{-2}(S(\varphi) + \gamma))$$

para todo  $\epsilon \leq \epsilon_0$ ,  $x \in \mathbb{R}^d$  e  $\varphi \in \Phi(s_0, x)$

(II)<sub>u</sub> Para cada  $\delta > 0$ ,  $\gamma > 0$  e  $s_0 > 0$  existe  $\epsilon_0 > 0$  tal que

$$P(\rho_T(X_\epsilon^x, \bar{\varphi}(s,x)) \geq \delta) \leq \exp(-\epsilon^{-2}(s-\gamma))$$

para todo  $s \leq s_0$ ,  $\epsilon \leq \epsilon_0$  e  $x \in \mathbb{R}^d$

Prova. (I)<sub>u</sub> e (II)<sub>u</sub> saem da prova do Teorema 2.3 desde que as  $B_x$  têm constante de Lipschitz que independe de  $x$ , e se  $\tilde{P}_\epsilon$  é a lei de  $(\epsilon W_t)_{0 \leq t \leq T}$  em  $C[0, T]$  então  $P_\epsilon^x = \tilde{P}_\epsilon \circ B_x^{-1}$ , por (2.1). Para (0)<sub>u</sub> observe que a transformação

$$h: \mathbb{R}^d \times \{\varphi \in C[0, T]: \varphi(0) = 0\} \longrightarrow \{\varphi \in C[0, T]: \varphi(0) = x\}$$

definida por  $h(x, \varphi) = B_x \varphi$  é um homeomorfismo. ■

### §2.2 Saída de um domínio

Seja  $D \subset \mathbb{R}^d$  um domínio limitado e cuja fronteira,  $\partial D$ , seja de classe  $C^2$  (para evitar technicalidades, mas não precisa tanto). A questão básica que agora discutiremos refere-se ao modo como as trajetórias de  $X_\epsilon^x$  saem de um tal domínio. Ou seja, se definimos

$$\begin{aligned} \tau_\epsilon^x &= \inf\{t > 0: X_\epsilon^x(t) \notin D\} \\ &= +\infty \quad \text{se tal conjunto for vazio} \end{aligned} \tag{2.21}$$

então perguntamos:

- (a) Que podemos dizer sobre  $X_\epsilon^x(\tau_\epsilon^x)$ ? (isto é, por onde  $X_\epsilon^x(\cdot)$  escapa de  $D$ ?)
- (b) Que sabemos sobre as v. aleatórias  $\tau_\epsilon^x$ ?

É bom contrastar pelo menos duas situações drasticamente diferentes: (i) a situação em que as trajetórias do sistema não perturbado  $X(t)$  sempre saem de  $\bar{D}$  em um tempo finito, conforme ilustrado na Figura 2.1. Neste caso a resposta a (a) e (b) nada tem a ver com grandes desvios; sabemos que para cada  $T > 0$  e  $\delta > 0$   $P(\rho_T(X_\epsilon, X) > \delta) \rightarrow 0$  e daí sai facilmente que existe  $t > 0$  tal que  $P(\tau_\epsilon^x \leq t) \rightarrow 1$ , e com probabilidade tendendo a um, a primeira saída ocorre próxima ao ponto por onde sai a trajetória correspondente do sistema determinístico, e em um tempo comparável; (ii) em  $\partial D$  o campo  $b(\cdot)$  aponta para dentro de  $D$  de modo que as trajetórias de  $X(\cdot)$  que iniciam em  $D$  nunca saem de  $D$  (i.e.,  $\langle b(x), n(x) \rangle < 0$  para todo  $x \in \partial D$ , onde  $n(x)$  denota o vetor unitário normal (exterior) a  $D$  em  $x$  (cf. Fig. 2.2.) Neste caso, a saída requer uma grande flutuação. Estamos então interessados no tempo necessário para a ocorrência de tal flutuação e no modo "mais provável" com que  $X_\epsilon$  sai de  $D$ . É nestas questões que estaremos nos concentrando.

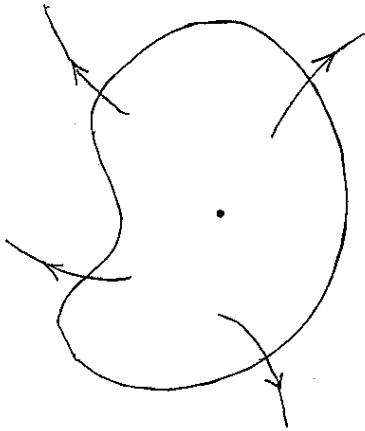


Fig. 2.1

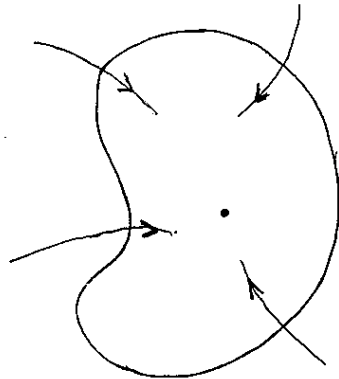


Fig. 2.2

Para simplificar nos situemos no caso particular em que todas as trajetórias de  $X(\cdot)$  iniciando em  $\bar{D}$  sejam atraídas a um único ponto de equilíbrio, assintoticamente estável  $x_0$ . Suponhamos então: ( $x_0 = 0$ )

- (a)  $D$  é domínio limitado  $\subseteq \mathbb{R}^d$  com  $\partial D$  de classe  $C^2$  e  $\langle b(x), n(x) \rangle < 0 \quad \forall x \in \partial D$ , onde  $n(x)$  é o vetor unitário normal (exterior) a  $D$ , em  $x$ .
  - (b)  $0 \in D$  é um ponto de equilíbrio estável (i.e. para toda  $U = \text{viz}(0)$ ,  $\exists V = \text{viz}(0)$  tal que  $0 \in V \subseteq U$  e se  $x \in V$  então  $X^x(t) \rightarrow 0$  sem sair de  $U$ );
  - (c)  $D$  é atraído a  $0$  i.e.,  $X^x(t) \rightarrow 0$  para todo  $x \in \bar{D}$ .
- (2.22)



Neste caso, conforme veremos, um papel importantíssimo no estudo das questões apresentadas será desempenhado pelo "quasi-potencial"  $V(0, x)$ , a seguir definido:

Definição 2.6.

$$V(0, x) = \inf\{S_{T_1, T_2}(\varphi) : \varphi(T_1) = 0, \quad \varphi(T_2) = x, \quad -\infty < T_1 < T_2 < +\infty\}$$

onde

$$S_{T_1, T_2}(\varphi) = \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(s) - b(\varphi(s))|^2 ds \quad \text{se } \varphi \text{ absolutamente cont\xed-nua;}$$

$$= +\infty, \quad \text{caso contr\xe1rio.}$$

De fato, um resultado que pode ser obtido com a utilização das estimativas de grandes desvios pode ser assim enunciado:

Teorema 2.7 (Freidlin e Wentzell). Suponhamos que  $b(\cdot)$  verifique as condições (2.2) e que as hipóteses (2.22) sejam satisfeitas. No caso de existir um único  $y_0 \in \partial D$  tal que

$$V(0, y_0) = \min_{y \in \partial D} V(0, y)$$

então, para todo  $\delta > 0$  e todo  $x \in D$

$$P_x(|X_{\epsilon}(\tau_{\epsilon}) - y_0| > \delta) \rightarrow 0 \tag{2.23}$$

onde

$$\tau_{\epsilon} = \inf\{t > 0 : X_{\epsilon}(t) \notin D\} \tag{2.24}$$

Ademais se  $K \subseteq D$  for um compacto, a convergência em (2.23) é uniforme sobre  $K$ .

Notações: (a)  $\tau_\epsilon(A)$  denotará o instante da primeira visita ao conjunto  $A$ , onde  $A$  é um boreliano de  $\mathbb{R}^d$ , i.e.

$$\tau_\epsilon(A) = \inf\{t > 0: X_\epsilon \in A\} \quad (2.25)$$

e usamos  $\tau_\epsilon = \tau_\epsilon(D^c)$ .

Estamos omitindo a notação referente ao ponto inicial pois ela estará indicada na probabilidade, e faremos isto com frequência. Também omitiremos o  $\epsilon$ , quando não causar confusão.

$$(b) \quad B_r(x) = \{y \in \mathbb{R}^d: |y-x| < r\}$$

Prova: Esta prova de Freidlin e Wentzell consiste nos dois pontos básicos seguintes

(i) Qualquer que seja o ponto inicial  $x$  e para qualquer vizinhança  $U$ , de  $O$ , a probabilidade de  $X_\epsilon$  escapar de  $D$  antes de visitar  $U$  tende a zero quando  $\epsilon \rightarrow 0$ . Pela propriedade de forte de Markov isto permite uma redução para pontos iniciais arbitrariamente próximos de  $O$ . Ou seja, basta que para cada  $\delta > 0$  possamos encontrar  $\mu > 0$  tal que  $\sup_{|y|=\mu} P_y(|X_\epsilon(\tau_\epsilon) - y_0| > \delta)$  convirja a zero. De fato, para cada  $x \in D$

$$\begin{aligned}
 P_x(|X_e(\tau_e) - y_0| > \delta) &= P_x(|X_e(\tau_e) - y_0| > \delta; \tau_e < \tau(B_\mu(0))) + \\
 &+ P_x(|X_e(\tau_e) - y_0| > \delta, \tau_e > \tau(B_\mu(0))) \leq \quad (2.26) \\
 &\leq P_x(\tau_e < \tau(B_\mu(0))) + \sup_{|y|=\mu} P_y(|X_e(\tau_e) - y_0| > \delta)
 \end{aligned}$$

onde na última passagem usamos a propriedade forte de Markov. É fácil ver que  $P_x(\tau_e < \tau(B_\mu(0))) \xrightarrow[\epsilon \downarrow 0]{} 0$  se  $B_\mu(0) \subset D$  e  $\mu > 0$ . Logo nossos esforços se concentrarão no segundo termo.

(ii) Dado  $\delta > 0$  é possível achar  $\mu > 0$  tal que

$$\sup_{|y|=\mu} P_y(|X_e(\tau_e) - y_0| > \delta) \rightarrow 0 \quad (2.27)$$

Para isto tomemos  $\mu > 0$  tal que  $B_{2\mu}(0) \subset D$ , definamos

$$\gamma = \{y: |y| = \mu\}$$

$$\Gamma = \{y: |y| = 2\mu\}$$

e consideremos uma Cadeia de Markov discreta imersa em  $X_e$  através dos tempos de parada  $\eta_n$  definidos a seguir:

$$\eta_0 = 0$$

$$\sigma_0 = \inf\{t > 0 : X_e(t) \in \Gamma\}$$

$$\eta_1 = \inf\{t > \sigma_0 : X_e(t) \in \gamma \cup \partial D\}$$

$$\sigma_1 = \inf\{t > \eta_1 : X_e(t) \in \Gamma\}$$

A cadeia de Markov auxiliar é  $Z_n^\epsilon = X_\epsilon(\eta_n)$ ,  $n \geq 1$ , com espaço de estados  $\gamma \cup D$ . Se  $\nu = \min\{n \geq 1: Z_n^\epsilon \in \partial D\}$ , então para  $y \in \gamma$

$$P_y(|X_\epsilon(\tau_\epsilon) - y_0| > \delta) = P_y(Z_\nu^\epsilon - y_0 | > \delta)$$

e queremos  $\mu$  de modo que isto tenda a zero com  $\epsilon$ .

Note que para cada  $y \in \gamma$ ,  $P_y(Z_1 \in \partial D)$  tende a zero, (cf (i)) mas queremos  $\mu > 0$  tal que

$$\sup_{y \in \gamma} P_y(Z_1 \in \partial D \setminus B_\delta(y_0) | Z_1 \in \partial D) \longrightarrow 0 \quad (2.28)$$

ou seja, dado que  $Z^\epsilon$  atinge  $\partial D$  é muito provável que isto aconteça em  $B_\delta(y_0)$ . Que (2.28) implica (2.27), cf desejado, sai facilmente da propriedade forte de Markov. Com efeito

$$\begin{aligned} P_y(Z_\nu^\epsilon \in \partial D \setminus B_\delta(y_0)) &= \\ &= \sum_n P_y(\nu = n, Z_n^\epsilon \in \partial D \setminus B_\delta(y_0)) = \\ &= \sum_n E_y \left[ 1(\nu \geq n) P_{Z_{n-1}}(Z_1 \in \partial D \setminus B_\delta(y_0)) \right] \\ &\leq \sum_n E_y \left[ 1(\nu \geq n) \cdot P_{Z_{n-1}}(Z_1 \in \partial D) \frac{P_{Z_{n-1}}(Z_1 \in \partial D \setminus B_\delta(y_0))}{P_{Z_{n-1}}(Z_1 \in \partial D)} \right] \\ &\leq \sup_{z \in \gamma} P_z(Z_1 \in \partial D \setminus B_\delta(y_0) | Z_1 \in \partial D). \end{aligned}$$

O que precisamos então para mostrarmos a viabilidade do método é a verificação de (2.28), para certo  $\mu > 0$  (suficientemente pequeno). Antes disto uma observação com relação à Cadeia de Markov  $(Z_n^\epsilon)$ .

Observação: Em princípio podemos ter  $\eta_n(\omega) = +\infty$  (e então  $\eta_m(\omega) = +\infty$  para todo  $m \geq n$ ) o que nos impediria de definir  $Z_n^\epsilon(\omega)$  pela regra dada acima. A dificuldade pode ser facilmente evitada em pelo menos dois modos: desprezando um conjunto excepcional que tem probabilidade zero, a única maneira de termos algum  $\eta_m(\omega) = +\infty$  é quando  $X_{\eta_n}^\epsilon(\omega) \in \partial D$  o que permite a possibilidade  $\sigma_n(\omega) = +\infty$  e então  $\eta_{n+1}(\omega) = +\infty$ . Entretanto isto só pode ocorrer após o processo  $X_\epsilon$  sair de  $D \cup \partial D$ . Portanto uma solução seria modificar o campo de vetores  $b(\cdot)$  fora de  $D \cup \partial D$  de modo que o processo sempre retorne a  $\Gamma$  e então também a  $\gamma$ . Tal modificação, não alterando  $b(\cdot)$  em  $\bar{D}$  manterá iguais as probabilidades a que se refere o Teorema 2.7. Outra solução é simplesmente acrescentar um "cemitério"  $\delta$  ao espaço de estados  $\gamma \cup \partial D$ , e definir  $Z_n^\epsilon(\omega) = \delta$  se  $\eta_n(\omega) = +\infty$ . Como  $Z_n^\epsilon$  só poderá ir para  $\delta$  após passar por  $\partial D$  isto não altera as probabilidades dos eventos que nos interessam.

Voltando à prova de (2.28), queremos mostrar que dado  $\delta > 0$  podemos tomar  $\mu > 0$  tal que

$$\frac{\sup_{x \in \gamma} P_x(Z_1 \in \partial D \setminus B_\delta(y_0))}{\inf_{x \in \gamma} P_x(Z_1 \in \partial D)} \longrightarrow 0 \text{ quando } \epsilon \downarrow 0 \quad (2.28)'$$

Utilizaremos então dois lemas básicos:

Lema 2.8. Nas hipóteses do Teorema 2.7, dado  $\alpha > 0$  temos:

- (a) existem  $a > 0$  e  $T_0 > 0$  tais que

$S_T(\varphi) \geq a(T-T_0)$  sempre que  $\varphi(t) \in \bar{D} \setminus B_\alpha(0)$ ,  $0 \leq t \leq T$

para todo  $T > T_0$ .

(b) existem  $c > 0$ ,  $T_0 > 0$  tais que se, sendo

$$R_\alpha = \inf\{t: X_\epsilon(t) \in D^c \cup B_\alpha(0)\}$$

então para  $\epsilon$  suficientemente pequeno

$$P_x(R_\alpha > T) \leq \exp(-\epsilon^{-2} c(T-T_0))$$

para todo  $x \in \bar{D} \setminus B_\alpha(0)$ .

Observação. O que está por trás deste lema é que se um compacto  $C$ , tal como  $\bar{D} \setminus B_\alpha(0)$ , (i.e. com algumas condições de regularidade) não possui nenhum "ponto de equilíbrio" para o sistema determinístico (i.e.,  $b(x_0) = 0$ ) então para qualquer  $c > 0$  existe  $T < +\infty$  tal que a probabilidade de permanecer em  $C$  durante  $[0, T]$  é menor do que  $\exp(-\epsilon^{-2}c)$  desde que  $\epsilon$  seja suficientemente pequeno, independentemente do ponto inicial em  $C$ .

Prova do Lema 2.8: (a) Por (2.22), se definirmos:

$$T(\alpha, x) \stackrel{\text{def}}{=} \inf\{t > 0: X^x(t) \in B_{\alpha/2}(0)\}$$

então:

- (1)  $T(\alpha, x) < +\infty$  para cada  $x \in \bar{D}$
- (2)  $x \rightarrow T(\alpha, x)$  é semicontínua superiormente pois  $X^x(t)$  depende continuamente de  $x$ .

Conseqüentemente  $T_0 \stackrel{\text{def}}{=} \sup_{x \in \bar{D}} T(\alpha, x) < +\infty$ , e se definimos

$$F = \{\varphi \in C[0, T_0] : \varphi(t) \in \bar{D} \setminus B_\alpha(0) \text{ para todo } t \in [0, T_0]\},$$

então  $F$  é fechado em  $C[0, T_0]$  e não contém nenhuma trajetória do sistema determinístico  $X(\cdot)$ , de modo que  $A \stackrel{\text{def}}{=} \min_{\varphi \in F} S_{T_0}(\varphi) > 0$ .

(Exercício 2.3). Seja  $T > 0$  e suponhamos  $\varphi(t) \in \bar{D} \setminus B_\alpha(0)$  para todo  $t \in [0, T]$ . Então, se  $T \geq T_0$

$$S_T(\varphi) \geq S_{T_0}(\varphi) \geq A.$$

Se  $T \geq 2T_0$  a aditividade de  $S_\cdot(\varphi)$  nos dá  $S_T(\varphi) \geq 2A$ , etc... de modo que para todo  $T \geq T_0$

$$S_T(\varphi) \geq A \lfloor T/T_0 \rfloor \geq A \left( \frac{T}{T_0} - 1 \right) = a(T - T_0)$$

onde  $a = A/T_0$ , provando (a).

(b) Seja  $\eta \in (0, \alpha/2)$  suficientemente pequeno de modo que para o domínio  $D^\eta \stackrel{\text{def}}{=} \{y : \text{distância}(y, D) < \eta\}$  as condições

(2.22) estejam verificadas. Então, por (a) existem  $A, T_0$  positivos tais que  $\varphi([0, T_0]) \subset D^\eta \setminus B_{\alpha/2}(0)$  implica  $S_{T_0}(\varphi) > A$ .

Logo se  $S_{T_0}(\varphi) \leq A$  e  $\psi([0, T]) \subset \bar{D} \setminus B_\alpha(0)$  temos  $\rho_{T_0}(\varphi, \psi) \geq \eta$ .

Daí vem que dado  $\gamma > 0$  existe  $\epsilon_0$  tal que

$$P_x(R_\alpha > T_0) \leq P_x(\rho_{T_0}(X_\epsilon, \Phi(x, A)) \geq \eta) \leq \exp(-\epsilon^{-2}(A - \gamma))$$

para todo  $x$ , desde que  $\epsilon \leq \epsilon_0$ . Pela propriedade de Markov obtemos

$$P_x(R_\alpha > (n+1) T_0) \leq P_x(R_\alpha > nT_0) \sup_y P_y(R_\alpha > T_0)$$

de modo que, por indução, para todo  $x \in \bar{D} \setminus B_\alpha(0)$ :

$$P_x(\tau_\alpha > T) \leq \exp - \epsilon^{-2} \left( \frac{T}{T_0} - 1 \right) (A - \nu)$$

se  $\epsilon \leq \epsilon_0$ , o que prova (b) com  $c = \frac{A-\nu}{T_0}$ . ■

Lema 2.9. Se o campo de vetores  $b(\cdot)$  verifica (2.2) então existe  $L < +\infty$  tal que para quaisquer  $x, y \in \mathbb{R}^d$  existe  $\varphi$  (de classe  $C^\infty$ ) com  $\varphi(0) = x$ ,  $\varphi(|x-y|) = y$  e  $S_{|x-y|}(\varphi) \leq L|x-y|$ .

Prova do Lema 2.9. Basta tomar  $\varphi(t)$ ,  $0 \leq t \leq |x-y|$  descrevendo o segmento que une  $x$  a  $y$ , i.e.,

$$\varphi(t) = x + \frac{t}{|y-x|} \cdot (y-x).$$

Então  $\varphi(0) = x$ ,  $\varphi(|x-y|) = y$  e

$$\begin{aligned} S_{|x-y|}(\varphi) &= \frac{|y-x|}{2} \left| \frac{y-x}{|y-x|} - b\left(\frac{y-x}{|y-x|}\right) \right|^2 \leq \\ &\leq L|y-x| \end{aligned}$$

onde  $L = \frac{1}{2} \sup_{|Z|=1} (|Z| + |b(Z)|)^2$ . ■

Conclusão da prova do Teorema 2.7. Para provarmos (2.28)', utilizaremos a seguinte escolha de  $\mu$ :

seja  $d = \min_{y \in \partial D \setminus B_\delta(y_0)} V(0, y) - V(0, y_0)$ , de modo que  $d > 0$



por hipótese; tomamos  $0 < \mu < \frac{d}{20L}$  onde  $L$  é dado pelo Lema 2.9.

Mostraremos que para tal escolha de  $\mu$ , existe  $\epsilon_0 > 0$  tal que se  $\epsilon \leq \epsilon_0$ :

$$P_x(Z_1 \in \partial D) \geq \exp\{-\epsilon^{-2}(V(0, y_0) + 0.4d)\} \quad (2.29)$$

$$P_x(Z_1 \in \partial D \setminus B_\delta(y_0)) \leq \exp\{-\epsilon^{-2}(V(0, y_0) + 0.5d)\} \quad (2.30)$$

para todo  $x \in \gamma$ .

Tendo em vista (2.29), verificaremos que para  $0 < \mu < d/10L$  pode-se garantir a existência de  $T > 0$  de modo que para  $x \in \gamma$  exista  $\varphi^x \in C[0, T]$  com  $\varphi^x(0) = x$ ,  $S_T(\varphi^x) \leq V(0, y_0) + 0.3d$  e tal que se  $\psi \in C[0, T]$  e  $\rho_T(\psi, \varphi^x) \leq \mu$  então após bater em  $\Gamma$ ,  $\psi(\cdot)$  sai de  $D$  antes de retornar a  $\gamma$ . Aliando-se isto à condição na definição de funcional de ação obtemos (2.29). Para efetuarmos tal construção fixemos um ponto  $y_1 \notin \bar{D}$  com  $|y_1 - y_0| = \mu$ , que servirá como ponto terminal das funções  $\varphi^x$ . A definição de  $V(0, y)$  nos permite encontrar  $T_1 > 0$  e  $\varphi^{(1)} \in C[0, T_1]$  tal que  $\varphi^{(1)}(0) = 0$ ,  $\varphi^{(1)}(T_1) = y_0$  e  $S_{T_1}(\varphi^{(1)}) \leq V(0, y_0) + d/20$ . Sejam  $t_1 = \sup\{t < T_1 : \varphi^{(1)}(t) \in \Gamma\}$  e  $x_1 = \varphi^{(1)}(t_1)$ .

Pelo Lema 2.9 podemos construir as seguintes funções:

- (i)  $\varphi^{(2)}(t): 0 \leq t \leq 2\mu$  tal que  $\varphi^{(2)}(0) = 0$ ,  $\varphi^{(2)}(2\mu) = x_1$  e  $S_{2\mu}(\varphi^{(2)}) \leq L \cdot 2\mu \leq d/10$ .

(ii)  $\varphi^{(3)}(t)$ :  $0 \leq t \leq \mu$  tal que  $\varphi^{(3)}(0) = x$ ,  $\varphi^{(3)}(\mu) = 0$  e  $S_\mu(\varphi^{(3)}) \leq L\mu \leq d/20$ .

(iii)  $\varphi^{(4)}(t)$ :  $0 \leq t \leq \mu$  tal que  $\varphi^{(4)}(0) = y_0$ ,  $\varphi^{(4)}(\mu) = y_1$  e  $S_\mu(\varphi^{(4)}) \leq L\mu \leq d/20$ .

Definimos então

$$\begin{aligned} \varphi^x(t) &= \varphi^{(3)}(t) && \text{se } 0 \leq t \leq \mu \\ &= \varphi^{(2)}(t-\mu) && \text{se } \mu \leq t \leq 3\mu \\ &= \varphi^{(1)}(t-3\mu+t_1) && \text{se } 3\mu \leq t \leq T_1-t_1+3\mu \\ &= \varphi^{(4)}(t-(T_1-t_1+3\mu)) && \text{se } T_1-t_1+3\mu \leq t \leq T_1-t_1+4\mu. \end{aligned}$$

Então  $\varphi^x(0) = x$ ,  $\varphi^x(T) = y$ , onde  $T = T_1 - t_1 + 4\mu$  ( $T$  independe de  $x$ ),  $S_T(\varphi^x) \leq V(0, y_0) + 0.3d$  e por construção  $\varphi^x$  verifica as condições exigidas, de modo que  $\rho_T(X_\epsilon, \varphi^x) < \mu$  implica  $Z_1 \in \partial D$ .

Logo, pela condição (I): para  $\epsilon$  suficientemente pequeno

$$P_x(Z_1 \in \partial D) \geq P_x(\rho_T(X_\epsilon, \varphi^x) < \mu) \geq \exp\{-\epsilon^{-2}(V(0, y_0) + 0.4d)\},$$

para todo  $x \in \gamma$ .

Voltemo-nos agora para (2.30). Pela propriedade forte de Markov temos: (omitindo  $\epsilon$  em  $\tau_\epsilon(\gamma \cup \partial D)$ , cf. (2.25))

$$\begin{aligned}
 P_x[Z_1 \in \partial D \setminus B_\delta(y_0)] &= P_x[X_\varepsilon(\eta_1) \in \partial D \setminus B_\delta(y_0)] = \\
 &= \int_\Gamma P_x[X_\varepsilon(\sigma_0) \in dy] P_y[X_\varepsilon(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0)] \leq \\
 &\leq \sup_{y \in \Gamma} P_y[X_\varepsilon(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0)] \quad (2.23)
 \end{aligned}$$

para todo  $x \in \gamma$ . Agora, para cada  $t > 0$

$$\begin{aligned}
 &P_y[X_\varepsilon(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0)] = \\
 &= P_y[X_\varepsilon(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0), \tau(\gamma \cup \partial D) \leq t] + \\
 &+ P_y[X_\varepsilon(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0), \tau(\gamma \cup \partial D) > t].
 \end{aligned} \quad (2.32)$$

Pelo Lema 2.8(b), dado  $c > 0$  posso tomar  $t_0$  tal que

$$P_y(\tau_\varepsilon(\gamma \cup \partial D) > t_0) \leq \exp(-c\varepsilon^{-2}) \quad \text{para todo } \varepsilon \leq \varepsilon_0,$$

todo  $y \in \Gamma$ ; apliquemos isto para  $c = V(0, y_0) + d$ . Resta estimar o primeiro termo no lado direito de (2.32). Note que se

$$F = \{\varphi \in C[0, t_0], \varphi(0) \in \Gamma \text{ e } S_{t_0}(\varphi) \leq V(0, y_0) + 0.6d\} \quad \text{então}$$

para toda  $\varphi \in F$ , e  $t \in [0, t_0]$ , distância  $(\varphi(t), \partial D \setminus B_\delta(y_0)) > \mu$ .

Com efeito, suponhamos que tal não ocorre e seja  $\varphi \in F$  e  $t_1$  tal que  $|\varphi(t_1) - z| \leq \mu$  para certo  $z \in \partial D \setminus B_\delta(y_0)$ , e  $t_1 \leq t_0$ .

Tomemos, pelo Lema 2.9:

$$(i) \quad \psi \in C[0, 2\mu] \quad \text{com } \psi(0) = 0, \quad \psi(2\mu) = \varphi(0) \quad e$$

$$S_{2\mu}(\varphi) \leq L \cdot 2\mu \leq \frac{d}{10}$$

(ii)  $\tilde{\psi} \in C[0, t_2]$  com  $t_2 \leq \mu$  tal que  $\tilde{\psi}(0) = \varphi(t_1)$ ,  
 $\tilde{\psi}(t_2) = z$  e  $S_{t_2}(\tilde{\psi}) \leq L\mu \leq d/20$ .

Definimos então

$$\begin{aligned} \tilde{\varphi}(t) &= \psi(t) && \text{se } 0 \leq t \leq 2\mu \\ &= \varphi(t-2\mu), && \text{se } 2\mu \leq t \leq 2\mu + t_1 \\ &= \tilde{\psi}(t-(2\mu+t_1)) && \text{se } 2\mu + t_1 \leq t \leq 2\mu + t_1 + t_2 = t_3 \end{aligned}$$

Então  $\tilde{\varphi} \in C[0, t_3]$ ,  $\tilde{\varphi}(0) = 0$ ,  $\tilde{\varphi}(t_3) = z \in \partial D \setminus B_\delta(y_0)$  e  
 $S_{t_3}(\tilde{\varphi}) \leq V(0, y_0) + 3d/4$  o que é absurdo pela escolha de  $d$ . Concluímos disto, pela condição (II) da Definição 2.1:

$$\begin{aligned} \sup_{y \in \Gamma} P_y(\tau(\gamma \cup \partial D) \leq t_0, X_e(\tau(\gamma \cup \partial D)) \in \partial D \setminus B_\delta(y_0)) \\ \leq \sup_{y \in \Gamma} P_y(\rho_T(X_e, F) \geq \mu) \\ \leq \exp\{-\epsilon^{-2}(V(0, y_0) + 0.55d)\} \end{aligned}$$

desde que  $\epsilon$  seja suficientemente pequeno. Daí obtemos (2.30) e então a prova do Teorema 2.7. ■

Neste ponto talvez sejam convenientes alguns comentários sobre possíveis generalizações do Teorema 2.7: (1) A mesma demonstração permite tratarmos o caso em que ao invés de atraído a um ponto de equilíbrio estável, o domínio  $D$  seja atraído a um ciclo limite estável, pelo sistema determinístico.

(2) Seria importante que pudéssemos ainda tratar situações em que  $D$  contém vários pontos de equilíbrio estável para o sistema não perturbado. (Neste caso deve-se esperar que a distribuição de saída do domínio  $D$  dependa da posição inicial.) Freidlin e Wentzell (1984) cuidaram também destas extensões do Teorema 2.7. Nestas notas estaremos omitindo a demonstração bem como uma discussão detalhada deste resultado, a fim de não nos estendermos demais. Entretanto, como no Capítulo 3 precisaremos utilizar tal resultado (em caso bem particular), recomendamos que na ocasião, o leitor interessado nos aspectos técnicos consulte [F-W (1984)].

Nos concentraremos agora na identificação de  $V(0, x)$  para uma classe de exemplos verificando (2.22); isto é importante nas aplicações do Teorema 2.7. Há um caso em que esta identificação é bem simples: quando o campo de vetores  $b(\cdot)$  deriva de um potencial

$$b(x) = - \nabla a(x)$$

onde  $a: \mathbb{R}^d \rightarrow \mathbb{R}$  continuamente diferenciável ( $C^2$  no nosso caso). De modo mais preciso, neste caso podemos verificar o seguinte:

Teorema 2.10. Suponhamos que  $b(\cdot)$  e  $D$  verifiquem as hipóteses do Teorema 2.7 e que exista  $a: \mathbb{R}^d \rightarrow \mathbb{R}$  tal que  $b(x) = -\nabla a(x)$ ,  $x \in \mathbb{R}^d$ . Então, para  $y \in \bar{D}$ :

$$V(0, y) = 2(a(y) - a(0)) \quad (2.34)$$

e para cada  $T > -\infty$ ,  $\inf\{S_{-\infty, T}(\varphi) : \varphi(-\infty) = 0, \varphi(T) = y\}$  é atin-

gido em uma única função  $\psi(\cdot)$  dada por

$$\psi(t) = X^y(T-t), \quad t \leq T \quad (2.35)$$

(i.e., a órbita do sistema determinístico que passa por  $y$ , com o sentido do tempo invertido).

Prova: Se  $\varphi \in C[T_1, T_2]$  e  $S_{T_1, T_2}(\varphi) < +\infty$  temos:

(2.36)

$$\begin{aligned} S_{T_1, T_2}(\varphi) &= \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(s) - b(\varphi(s))|^2 ds = \\ &= \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(s) + b(\varphi(s))|^2 ds - 2 \int_{T_1}^{T_2} \langle \dot{\varphi}(s), b(\varphi(s)) \rangle ds \\ &= \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(s) + b(\varphi(s))|^2 ds + 2 \int_{T_1}^{T_2} \langle \nabla a(\varphi(s)), \dot{\varphi}(s) \rangle ds \end{aligned}$$

onde usamos a identidade  $|u+v|^2 = |u-v|^2 + 4\langle u, v \rangle$ ,  $u, v \in \mathbb{R}^d$  e a hipótese  $b(x) = -\nabla a(x)$  em  $\mathbb{R}^d$ .

Segue-se de (2.36) que

$$S_{T_1, T_2}(\varphi) \geq 2(a(\varphi(T_2)) - a(\varphi(T_1)))$$

de modo que  $V(0, y) \geq 2(a(y) - a(0))$  para todo  $y \in \mathbb{R}^d$ . Por outro lado, se  $y \in \bar{D}$  a função  $\psi(\cdot)$  dada por (2.35) satisfaz a equação

$$\begin{aligned} \dot{\psi}(s) &= -b(\psi(s)) & s \leq T \\ \psi(T) &= y. \end{aligned} \tag{2.37}$$

sendo que  $\psi(-\infty) = 0$  e então por (2.36):

$$S_{-\infty, T}(\psi) = 2(a(y) - a(0)) \quad \text{e} \quad V(0, y) = 2(a(y) - a(0))$$

se  $y \in \bar{D}$ .

Por outro lado (2.36) garante que  $S_{T_1, T_2}(\varphi) = 2(a(\varphi(T_2)) - a(\varphi(T_1)))$  se e só se  $\varphi$  for absolutamente contínua em  $(T_1, T_2)$  e  $\int_{T_1}^{T_2} |\dot{\varphi}(s) + b(\varphi(s))|^2 ds = 0$ . Como a condição (2.2) garante unicidade da solução de

$$\dot{\psi}(t) = -b(\psi(t)), \quad t < T$$

$$\psi(T) = y$$

(cuja solução é  $\psi(t) = X^y(T-t)$ )

decorre a unicidade da função extremal no sentido acima descrito. ■

Observações: (1) Para (2.34) não utilizamos toda a força da hipótese (2.2) feita inicialmente sobre  $b(\cdot)$ . Com efeito, para (2.36) basta que  $b(x) = -\nabla a(x)$  para todo  $x \in \mathbb{R}^d$ , com  $a(\cdot)$  continuamente diferenciável. Se, ademais,  $a(\cdot)$  verifica  $a(0) < a(y)$  e  $\nabla a(y) \neq 0$  para todo  $y \in \bar{D} \setminus \{0\}$  então existe solução de

$$\dot{\psi}(t) = -b(\psi(t)).$$

$$\psi(T) = y$$

desde que  $y \in \bar{D}$ . Segue-se então que  $V(0,y) = 2(a(y) - a(0))$  para todo  $y \in \bar{D}$ . Quanto à unicidade da função extremal basta unicidade da solução de (2.37), o que ocorre se  $a$  for também de classe  $C^2$ .

(2) Outra pequena observação referente a (2.34): se estivermos supondo  $b(x) = -\nabla a(x)$  apenas para  $x \in \bar{D}$ , (2.34) continua válida desde que  $y \in \bar{D}$  e  $a(y) \leq \min_{z \in \partial D} a(z)$ . (Ver Exercício 2.8)

Assim, na situação descrita no Teorema 2.10  $V(0,x)$  corresponde a um múltiplo da diferença de potencial entre  $x$  e  $0$ . É esta uma das razões para que no caso geral  $V(x,y)$  seja chamado de "quase-potencial". Com relação ao caso geral, i.e. em que  $b(\cdot)$  não precisa ser o gradiente de uma função, podemos dizer que a função  $V(0,\cdot)$  é sempre contínua (mesmo Lipschitz) mas pode não ser diferenciável (Exercício 2.10). No caso de diferenciabilidade poderíamos utilizar a equação de Jacobi para o problema variacional que define  $V(0,x)$  e com isto tentar obter mais informações sobre  $V(0,\cdot)$ , de modo análogo ao caso tratado no Teorema 2.10. Freidlin e Wentzell provam que  $V(0,\cdot)$  será continuamente diferenciável se e só se  $b(x)$  se decompõe como

$$b(x) = -\nabla a(x) + \iota(x) \tag{2.38}$$



com  $a(\cdot)$  como na observação acima, e  $\langle t(x), \nabla a(x) \rangle = 0$  em  $\bar{D}$ . Neste caso temos em  $\bar{D}$

$$\frac{1}{2} |\nabla V(0,x)|^2 = - \langle b(x), \nabla V(0,x) \rangle \quad (2.39)$$

e o Teorema 2.10 tem uma extensão natural (F-W (1984) p. 119). Neste curso preferimos, por simplicidade, ficar apenas com o caso gradiente.

Um outro ponto do Teorema 2.10 merece comentários: é o ponto referente à unicidade da função  $\psi$  tal que  $\inf\{S_{-\infty, T}(\varphi); \varphi(-\infty) = 0, \varphi(T) = y\} = S_{-\infty, T}(\psi)$  para cada  $y \in D \cup \partial D$ ; ao variarmos  $T$ , a função  $\psi$  muda apenas por uma translação do tempo, e  $\psi$  corresponde à trajetória do sistema determinístico revertendo-se o sinal do tempo. Nesta situação o Teorema 2.10 pode ser refinado dizendo-se que não apenas  $X_{\epsilon}(\tau_{\epsilon})$  estará com grande probabilidade próximo a  $y_0$ , mas também o "segmento terminal" de  $X_{\epsilon}(\cdot)$  antes de sair de  $D$  estará muito próximo da curva onde  $S$  atinge seu mínimo  $V(0, y_0)$  (há sempre uma translação do tempo envolvida). O enunciado preciso (e mais geral) disto está no Capítulo 4 de F-W (1984), mas podemos resumí-lo assim no caso do Teorema 2.10: Se  $\psi(t) = X^{y_0}(-t)$ ,  $t \geq 0$  e  $U$  for vizinhança de  $0$ , contida em  $D$  com  $\partial U$  de classe  $C^2$  e tal que  $\psi$  sai de  $\partial U$  de forma regular, então para todo  $\delta > 0$

$$P\left(\sup_{\theta_{\epsilon} < t < \tau_{\epsilon}} |X_{\epsilon}(t) - \psi(t - \theta_{\epsilon} + \theta)| > \delta\right) \rightarrow 0$$

onde

$$\theta_\epsilon = \sup\{t < \tau_\epsilon : X_\epsilon(t) \in U\}$$

$$\theta = \sup\{t : \psi(t) \in U\}.$$

[Dizer que  $\psi$  sai de  $U$  de forma regular significa que existe  $\eta > 0$  tal que  $\psi(s) \in U$  para  $\theta - \eta < s < \theta$ , i.e. não oscila infinitamente entre  $U$  e  $\mathbb{R}^d \setminus U$ .]

A prova é baseada em argumentos análogos aos do Teorema 2.7, porém será omitida (F-W (1984), Teorema 2.3, Capítulo 4).

Vimos a importância do quase-potencial  $V(0, y)$  no estudo do comportamento assintótico de  $X_\epsilon(\tau_\epsilon)$  quando  $\epsilon \downarrow 0$ . Outros problemas relacionados referem-se

- (a) ao tamanho de  $\tau_\epsilon$ , i.e., estimativas para a média ou mediana de  $\tau_\epsilon$ ;
- (b) à lei de  $\tau_\epsilon$ , e.g.  $\tau_\epsilon / E\tau_\epsilon$  converge em lei?

Com relação a (a) Freidlin e Wentzell obtiveram os seguintes resultados: Nas condições do Teorema 2.7 e se

$V_0 \stackrel{\text{def}}{=} V(0, y_0) = \min_{y \in \partial D} V(0, y)$  então se  $x \in D$  e  $\eta > 0$  existe  $\epsilon_0 > 0$  tal que

$$\exp\left(\frac{V_0 - \eta}{\epsilon^2}\right) \leq E_x \tau_\epsilon \leq \exp\left(\frac{V_0 + \eta}{\epsilon^2}\right) \quad (2.40)$$

para  $0 < \epsilon < \epsilon_0$ ; ademais:

$$P_x \left( e^{-\frac{V_0 - \eta}{\epsilon^2}} < \tau_\epsilon < e^{\frac{V_0 + \eta}{\epsilon^2}} \right) \xrightarrow{\epsilon \downarrow 0} 1 \quad (2.41)$$

Pretendemos pois dar uma maior atenção ao ítem (b) o que será feito no Capítulo 3 (e onde utilizaremos estes resultados de Freidlin e Wentzell que acabamos de mencionar, mas que não os provaremos neste texto. Nos exercícios há indicações da prova, que o leitor poderá completar. (F-W (1984), Capítulo 4).

Por enquanto nos limitaremos a uma descrição heurística do que se espera em (b). Os resultados precisos serão vistos no Capítulo 3.

Ao observarmos  $X_\epsilon(\cdot)$  durante um intervalo de tempo de comprimento fixado e.g.  $[k, k+1)$  é muito provável que vejamos uma trajetória bem próxima<sup>(1)</sup> à do sistema determinístico (não perturbado)  $X(\cdot)$ , com mesma condição inicial. Por outro lado, as condições do Teorema 2.7 garantem que para  $k$  suficientemente grande  $X(t)$  esteja muito próximo ao ponto  $x_0$ , que atrai  $\bar{D}$  e que s.p.g. supusemos  $x_0 = 0 \in D$ . Estes dois fatos nos levam a acreditar que, definindo

$$A_k^\epsilon = [X_\epsilon(t) \notin D \text{ para certo } t \in [k, k+1)]$$

então  $P(A_k^\epsilon)$  não depende muito de  $k$ , desde que  $k$  seja suficientemente grande, e converge a zero quando  $\epsilon \downarrow 0$ , i.e. existem  $p_\epsilon > 0$  com  $p_\epsilon \downarrow 0$  quando  $\epsilon \downarrow 0$  tais que  $P(A_k^\epsilon) \approx p_\epsilon$  desde que  $k$  seja suficientemente grande. A razão heurística fica clara ao escrevermos  $P(A_k^\epsilon) = \int P_y(A_0^\epsilon) P(X_\epsilon(k) \in dy)$ , pela propriedade de Markov, e ao lembrarmos que quando  $k$  é suficientemente grande  $P(X_\epsilon(k) \in dy)$  concentra-se cada vez mais em torno a zero. Além disto se  $k \ll k'$  e  $k'-k$  grande,  $(X_\epsilon(t), t \in [k, k+1))$

(1) Próxima no sentido da norma do supremo.

e  $(X_\epsilon(t): t \in [k', k'+1])$  são "quase independentes" (o mesmo para um número finito qualquer de  $k$ 's. Isto pode ser justificado pela propriedade de Markov e porque a lei de  $X_\epsilon(s)$  não depende muito do ponto inicial  $X_\epsilon(0)$ , quando  $s$  for grande, pelas observações acima.

Estes argumentos heurísticos sugerem que a variável aleatória  $[\tau_\epsilon]$  (parte inteira de  $\tau_\epsilon$ ) deve ter uma distribuição aproximadamente geométrica com parâmetro  $p_\epsilon$ , quando  $\epsilon \downarrow 0$ , pois:

$$[\tau_\epsilon] = \min\{n \geq 1: w \in A_n^\epsilon\},$$

(i.e., a descrição heurística sugere que  $[\tau_\epsilon]$  deva se comportar, quando  $\epsilon \downarrow 0$ , como o número de lançamentos necessários até obtenção de uma cara, quando os lançamentos são independentes e em cada lançamento a probabilidade de obter cara seja  $p_\epsilon$ .

No Capítulo 3 daremos um argumento rigoroso, mostrando que existe  $(\beta_\epsilon)$ ,  $\beta_\epsilon \nearrow +\infty$ , de modo que  $\tau_\epsilon/\beta_\epsilon$  convergem em lei a uma exponencial com média 1 i.e.  $P_y(\tau_\epsilon/\beta_\epsilon > t) \rightarrow e^{-t}$  quando  $\epsilon \downarrow 0$ , para cada  $y \in D$ .

Este resultado, nas condições do Teorema 2.7, foi provado por M. Day, (usando métodos probabilísticos e analíticos), sendo que  $\beta_\epsilon = E_0 \tau_\epsilon$ . No Capítulo 3 necessitamos de um resultado análogo, mas como estamos interessados na passagem de  $X_\epsilon(\cdot)$  de uma bacia de atração para outra, veremos que as condições do Teorema 2.7 não se verificam, pois  $D$  possui mais de um ponto crítico para o sistema determinístico. Veremos então uma outra prova, puramente probabilística, e que aparece em (Galves, Olivieri, e Vares (1984)). Está feita no caso de  $b$  ser um campo:

gradiente, mas vale em condições mais gerais, que podem ser colocadas de modo preciso.

### Exercícios

2.1 (a) Seja  $(B_t)_{t \geq 0}$  um movimento Browniano real com  $B_0 = 0$  e coeficiente de difusão igual a um (i.e. o m. Browniano padrão começando em zero). O "princípio da reflexão" permite calcular a distribuição conjunta de  $M_t = \max_{s \leq t} B_s$  e  $B_t$  para cada  $t > 0$ , através da relação

$$P(M_t > b, B_t < b-a) = P(B_t > b+a) \quad (*)$$

se  $a, b > 0$ . (Para a prova de (\*) referimos a algum dos textos básicos tais como Breiman (1968), ou Williams (1979).)

Deduza de (\*) que

$$P(M_t > b) = 2P(B_t > b)$$

para todo  $b > 0$ .

(Aconselho uma figura para entender o conteúdo de (\*).)

(b) Verifique que

$$\int_x^{+\infty} e^{-u^2/2} du \leq \frac{1}{x} e^{-x^2/2}$$

para todo  $x > 0$ . (Sugestão:  $\int_x^{+\infty} ue^{-u^2/2} du = e^{-x^2/2}$ )

(c) Utilize os ítems anteriores para deduzir as desigualdades (2.8) e (2.16).

## 2.2 (Comparação da Definição 2.1 com os métodos do Capítulo 1)

(a) Verifique que a condição (I) na Definição 2.1 equivale à seguinte:

(I') Se  $A$  aberto  $\subseteq M$

$$\lim_{\epsilon \downarrow 0} \lambda(\epsilon)^{-1} \ln \mu_{\epsilon}(A) \geq -\inf\{S(y) : y \in A\}$$

(b) Mostre que (O) e (II) na Definição 2.1 implicam a seguinte condição:

(II') Se  $A$  fechado  $\subseteq M$

$$\overline{\lim}_{\epsilon \downarrow 0} \lambda(\epsilon)^{-1} \ln \mu_{\epsilon}(A) \leq -\inf\{S(y) : y \in A\}.$$

(c) Mostre que (II') implica (II) na Definição 2.1. Conclua que a Definição 2.1 para um funcional de ação pode ser reescrita pelas condições (O), (I'), (II').

(Compare com as discussões no final do Capítulo 1).

(d) Tendo definido  $I(A) = \inf_{y \in A} S(y)$ , diremos que um boreliano  $A \subseteq M$  é regular se  $I(\bar{A}) = I(\overset{\circ}{A})$ , onde  $\bar{A}$  = fecho de  $A$  e  $\overset{\circ}{A}$  = interior de  $A$ . Conseqüentemente se  $\lambda(\epsilon)$ ,  $S(\cdot)$  verificam a Definição 2.1, temos

$$(I'') \quad \frac{1}{\lambda(\epsilon)} \lim_{\epsilon \downarrow 0} \mu_\epsilon(A) \xrightarrow{\epsilon \downarrow 0} -I(A)$$

para todo  $A$  regular. Prove que, reciprocamente, as condições (0) e (I'') implicam (I) e (II) i.e. garantem que  $\lambda(\epsilon)$ ,  $S(\cdot)$  satisfazem a definição de funcional de ação.

(e) Nas condições da Definição 2.1 verifique que para um bo-reliano  $A$  que seja regular e tal que exista um único  $x_0 \in A$  com  $S(x_0) = I(A)$  então

$$\frac{\mu_\epsilon(A \cap B_\delta(x_0))}{\mu_\epsilon(A)} \xrightarrow{\epsilon \downarrow 0} 1$$

para qualquer  $\delta > 0$  ( $B_\delta(x_0) = \{y: \rho(y, x_0) < \delta\}$ )

(d) O seguinte resultado (F-W (1984)) permite garantir a unicidade do funcional de ação  $S(\cdot)$  uma vez fixadas as constantes  $\lambda(\epsilon)$ . De fato, o leitor pode verificar:

$$\begin{aligned} -S(y) &= \lim_{\delta \downarrow 0} \overline{\lim}_{\epsilon \downarrow 0} \lambda(\epsilon)^{-1} \lim_{\epsilon \downarrow 0} \mu_\epsilon(B_\delta(y)) \\ &= \lim_{\delta \downarrow 0} \underline{\lim}_{\epsilon \downarrow 0} \lambda(\epsilon)^{-1} \lim_{\epsilon \downarrow 0} \mu_\epsilon(B_\delta(y)) \end{aligned}$$

(compare com a definição de  $-\lambda_\mu(x)$  no Capítulo 1.)

2.3 Verifique que sob a condição (0) da Definição 2.1 o funcional  $S(\cdot)$  atinge o mínimo sobre qualquer conjunto fechado.

2.4 Mostre que se  $M$  for espaço métrico completo, a condição (0) na Definição 2.1 equivale a (0.a) e (0.b), onde

(0.a)  $\Phi(s)$  é relativamente compacto;

(0.b)  $\Phi(s)$  é fechado;

e (0-b) é equivalente à semicontinuidade inferior de  $S(\cdot)$ .

Note que esta última condição não é realmente uma restrição, pois se  $\lambda(\epsilon)$ ,  $S(\cdot)$  verificam (0-a), (I) e (II) e substituímos  $S$  por  $\underline{S}$ , onde

$$S(y) = S(y) \wedge \liminf_{x \rightarrow y} S(x), \quad y \in M,$$

então  $\lambda(\epsilon)$ ,  $\underline{S}(\cdot)$  satisfazem a Definição 2.1.

2.5 Sendo  $X$  uma v.a Normal com média 0 e variância 1, calcule  $E e^{\theta Z^2}$ , com  $\theta > 0$ .

2.6 Conclua que  $\epsilon^{-2} S_T(\cdot)$  no Teorema 2.2 é um funcional de ação para  $(\epsilon W_t)$ , verificando que  $S_T(\cdot)$  definida por (2.13) é semicontínua inferiormente e que  $\Phi_T(s) \stackrel{\text{def}}{=} \{\varphi: S_T(\varphi) \leq s\}$  é relativamente compacto, para todo  $0 < s < +\infty$ . (Sugestão: Arzelà-Ascoli).

2.7 Prove o Teorema 2.3. (É simples rotina.)

2.8 Verifique que (2.34) continua válido desde que  $b$  satisfaça (2.2), (2.22) com  $b(x) = -\nabla a(x)$  para  $x \in \bar{D}$ , e que o ponto  $y \in \bar{D}$  verifique



$$a(y) \leq \min_{z \in \partial D} a(z).$$

2.9 Mostre através de um exemplo que  $V(x, \cdot)$  pode deixar de ser diferenciável (F-W (1984)).

## CAPÍTULO 3

### METAESTABILIDADE E GRANDES DESVIOS

#### §3.1 Comentários Gerais

No Capítulo anterior vimos que estimativas para probabilidades de certos eventos raros (grandes desvios) podem fornecer informações relevantes sobre o comportamento típico de certos processos estocásticos. Os Teoremas 2.7 e 2.10 exemplificam bem isto, dando informação sobre como (por onde, mais precisamente)  $X_\epsilon(\cdot)$  escapa do domínio  $D$ , invariante para o sistema determinístico. Outra questão importante refere-se ao tempo necessário para tal saída: estando em  $D$  o processo é (rapidamente) atraído em direção a  $x_0 \in D$ . Entretanto devido ao ruído aleatório ( $\epsilon \dot{W}_t$ ) ele faz excursões que podem afastá-lo  $x_0$ . Durante um intervalo  $[0, T]$  fixado será muito raro que uma tal excursão o arraste para fora de  $D$ , e então pelas características de  $b(\cdot)$  e  $D$ ,  $X_\epsilon(\cdot)$  será - com grande probabilidade - novamente atraído em direção a  $x_0$ , e a situação tende a se repetir em  $[T, 2T]$ , etc., até que após muitas tentativas frustradas  $X_\epsilon(\cdot)$  consiga escapar de  $D$ . Esta análise bem heurística sugere que o instante da primeira saída de  $D$  se pareça com o número de lançamentos necessários até obtenção da primeira cara em sucessivos lançamentos independentes de uma moeda, em que a probabilidade de obter cara seja muito pequena, levando naturalmente à conjectura de

que o tempo de saída de  $D$  tenha distribuição aproximadamente exponencial i.e.  $P(\tau_{\epsilon}/\beta_{\epsilon} > t) \sim e^{-t}$  quando  $\epsilon \neq 0$  para  $\beta_{\epsilon} > 0$  convenientes. (Day (1983); Galves Olivieri, Vares (1984)).

Uma variável aleatória exponencial  $\tau$  caracteriza-se pela imprevisibilidade a ela associada: saber que  $\tau$  ainda não ocorreu em  $[0, T]$  em nada modifica a distribuição do tempo que ainda resta até  $\tau$ . Isto é expresso pela identidade  $P(\tau > t+s \mid \tau > t) = P(\tau > s)$ . A lei exponencial, e a geométrica, sua versão discreta, são típicas na ocorrência de grandes flutuações.

Pois bem, fenômenos naturais cuja ocorrência requer uma "grande flutuação" são relativamente comuns. Aparecem com frequência em situações de transição de fase estando intimamente ligados ao fenômeno da metaestabilidade.

Por exemplo, sob condições especiais é possível manter a água em estado líquido, mesmo abaixo de sua temperatura crítica ( $0^{\circ} \text{C}$ ) por um tempo muito longo. Tem-se a impressão de se tratar de um sistema em equilíbrio; subitamente porém, forma-se um pequeno cristal de gelo e ocorre uma rápida transição para o verdadeiro estado de equilíbrio (sólido, neste caso).

Este fenômeno físico é conhecido como "metaestabilidade" e muitas teorias a respeito tem sido propostas. Entretanto o problema de modelagem e de uma descrição matematicamente rigorosa permanece.

Modelando-se os sistemas em questão através de processos de Markov o fenômeno da metaestabilidade seria assim caracteri-

zado: mesmo possuindo apenas uma probabilidade invariante  $\mu_e$  o processo comporta-se por um tempo muito longo como se estivesse em equilíbrio, descrito por outra medida  $\mu_m$ . Finalmente ocorrerá uma transição, brusca e imprevisível, para o verdadeiro equilíbrio  $\mu_e$ .

A maior parte das teorias existentes baseia-se na chamada "evolução de ensembles", ou seja na evolução da lei  $\mu(t)$  do processo, ao longo do tempo  $t$ , quando é dada uma lei inicial  $\mu(0)$ . Aqui os esforços se concentram em encontrar boas prescrições para medidas iniciais  $\mu$  que correspondam aos estados metaestáveis. Como o nome indica estas prescrições devem ter como base o comportamento de  $t \rightarrow \mu(t)$ , e são basicamente: se  $\mu(0) = \mu$  a evolução das médias  $\int f d\mu(t)$  às correspondentes ao equilíbrio é lenta; para uma ampla classe de leis iniciais  $\mu(0) \neq \mu$  as médias  $\int f d\mu(t)$  evoluem rapidamente a  $\int f d\mu$ . A literatura é bem extensa e nos referimos a Penrose e Lebowitz (1979) para uma discussão mais ampla e para referências.

Recentemente M. Cassandro, A. Galves, E. Olivieri e M. E. Vares (1984) propuseram um enfoque diferente, com base em estatísticas obtidas ao longo de trajetórias típicas. Este enfoque "trajetória por trajetória" propõe que na caracterização de comportamento metaestável se utilizem médias temporais ao longo de trajetórias típicas. Diríamos então que o sistema exibe comportamento metaestável se médias temporais convenientes apresentarem o seguinte comportamento: permanecem estáveis por um tempo muito longo, sofrem uma brusca transição e estabilizam-se novamente.

Esta descrição, embora imprecisa, nos faz sentir o nítido contraste com o "comportamento hidrodinâmico", quando ocorre evolução suave e gradual (em escalas convenientes) através de um contínuum de estados de equilíbrio. Neste curso não discutiremos sobre comportamento hidrodinâmico; fica apenas registrada a referência (A. De Masi et al. (1983)), uma excelente resenha das recentes atividades de pesquisa neste tópico.

Neste capítulo pretendemos apresentar os pontos básicos do chamado "enfoque trajetória por trajetória", através principalmente da discussão de alguns exemplos estudados em (M. Cassandro et al (1984); A. Galves et al (1984)).

Neste curso, esta discussão tem como objetivo destacar o papel de "grandes desvios" que, de forma bem geral, são responsáveis pela transição de fase na situação metaestável, e a utilização de teorias de grandes desvios. Um exemplo é a Teoria de Freidlin e Wentzell, que no caso das difusões multidimensionais, foi estudada no Capítulo 2. Neste capítulo será aplicada a uma classe de exemplos apresentando comportamento metaestável.

Sobre a possível utilidade concreta da proposta acima (enfoque trajetória por trajetória) para o entendimento do fenômeno físico da Metaestabilidade, ainda não podemos dizer muito. Em grande parte isto dependerá do estudo - sob este ponto de vista - de modelos de interesse na Mecânica Estatística, como é o caso do modelo estocástico de Ising bidimensional. Para discussões sobre o problema da metaestabilidade neste modelo referimos a D. Capocaccia et al. (1974). Binder e H. Krumbhaar (1974) possuem resul

tados numéricos sugestivos.

Com relação à proposta contida no chamado "enfoque trajetória por trajetória", a afirmação feita anteriormente sobre o comportamento típico de certas médias temporais é colocada sob forma de um teorema de limite, onde estaríamos analisando uma família de sistemas semelhantes, convenientemente parametrizada,  $(Y_\epsilon)_\epsilon$ . A aparente estabilidade de médias temporais (convenientemente tomadas) seguida de uma transição brusca e imprevisível ao equilíbrio verdadeiro, traduzir-se-á pela convergência em lei de processos (a valor medida) definidos através de médias temporais convenientes, a um processo Markoviano de salto, que vale  $\mu_m$  antes do salto, e depois vale  $\mu_e$ . Aqui, a propriedade de Markov é uma característica importante; corresponde à lei exponencial para o instante de salto, e revela o caráter imprevisível da transição. A afirmação precisa está contida nos teoremas enunciados na seção seguinte.

### §3.2 Comportamento metaestável de uma classe de sistemas dinâmicos sujeitos a pequenas perturbações estocásticas

Utilizando mesma notação do Capítulo 2, consideremos uma família de processos de difusão  $X_\epsilon(\cdot)$  definidos pela equação (2.1), onde o campo de vetores  $b$  satisfaz (2.2) e, ademais, deriva de um potencial de dois poços, i.e.,

$$b(x) = -\nabla a(x) \quad x \in \mathbb{R}^d \quad (3.1)$$

onde  $a(\cdot)$  é da forma indicada pela Figura 3.1. De modo mais preciso:

- (i)  $a: \mathbb{R}^d \rightarrow \mathbb{R}$  é de classe  $C^2$ ;
- (ii)  $a(x) \rightarrow +\infty$  se  $|x| \rightarrow +\infty$ ;
- (iii)  $a(\cdot)$  tem exatamente três pontos críticos, denotados por  $p, q$  e  $r$ , todos hiperbólicos, i.e.,  
 $\det \left( \frac{\partial^2 a(x)}{\partial x_i \partial x_j} \right) \Big|_{x=p,q,r} \neq 0$ , sendo  $p$  e  $q$  pontos de equilíbrio estável de (2.3), e sendo  $r$  um ponto de sela;
- (iv)  $a(q) < a(p) < a(r)$

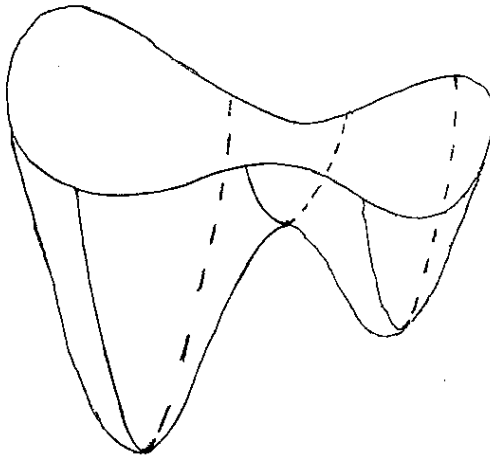


Fig. 3.1

Note que na inexistência do ruído, o comportamento do sistema dinâmico dado por

$$\dot{x}(t) = b(x(t)) \quad (3.3)$$

é muito simples: tem-se a decomposição  $\mathbb{R}^d = D_p \cup D_q \cup W_r^S$  onde  $D_p, D_q$  são domínios de  $\mathbb{R}^d$  tais que se  $x = X(0) \in D_p$  ( $D_q$ ) então  $X(t) \in D_p$  para todo  $t > 0$  ( $D_q$ , resp.) e  $X(t) \rightarrow p$  ( $q$  resp.),  $W_r^S = \bar{D}_p \cap \bar{D}_q$  tem a propriedade de que se  $x = X(0) \in W_r^S$  então  $X(t) \rightarrow r$ , sem nunca sair de  $W_r^S$ .

Observação:  $D_p, D_q$  são as bacias de atração dos atratores  $p$  e  $q$  respectivamente.  $W_r^S$  é uma separatriz da sela  $r$ , ou a variedade estável correspondente à sela  $r$ , e tem codimensão 1, sob as hipóteses dadas. A variedade instável de  $r$ ,  $W_r^U$  tem dimensão 1.

Agora, a presença da perturbação estocástica permite a passagem de  $D_p$  a  $D_q$  e vice-versa. Para fixar idéias supomos que o ponto inicial esteja em  $D_p$ ; então, o comportamento típico para  $X_\epsilon(\cdot)$  deve ser o seguinte:

(a) Inicialmente  $X_\epsilon(\cdot)$  é atraído em direção a  $p$ , devido à atuação do campo  $b(\cdot)$ , embora sofrendo perturbações "erráticas" devidas ao "ruído" ( $\epsilon \dot{W}_t$ ); estando muito próxima a  $p$  a importância desta perturbação tende a superar largamente a do campo (pois  $|b(x)| \leq K \delta_\epsilon$  se  $|x-p| \leq \delta_\epsilon$ ) e isto permite que a trajetória se afaste de  $p$ . Entretanto tende a ser atraída novamen



te e tudo parece recomeçar, com um número muito grande de tentativas para escapar de  $D_p$ , logo seguidas de nova atração (forte) em direção a  $p$ . Portanto enquanto perdurar esta situação o processo passa a maior parte do tempo próximo a  $p$ ; dá a impressão de um processo em equilíbrio com medida estacionária muito próxima a  $\delta_p$  quando  $\epsilon \downarrow 0$ . É isto que chamamos de "termalização".

(b) Após muitas tentativas frustradas  $X_\epsilon(\cdot)$  consegue vencer a barreira do campo e passar para  $D_q$ . De acordo com a intuição, e com os resultados do Capítulo 2, acreditamos que a passagem ocorra — com grande probabilidade — em lugar próximo a  $r$ . Ademais, o tempo decorrido até esta transição, lembra algo como o número de lançamentos necessários até obtenção de cara em sucessivos lançamentos independentes de uma moeda, quando a probabilidade de obter cara é muito pequena. Daí, este tempo deveria ter uma distribuição aproximadamente exponencial.

(c) Uma vez em  $D_q$  a coisa toda se repete, só que em relação ao ponto  $q$ . Entretanto como  $a(q) < a(p)$  o tempo para retornar a  $D_p$  será infinitamente maior que o anterior (quando  $\epsilon \downarrow 0$ ).

Ou seja, o comportamento típico das trajetórias parece caracterizar-se por oscilações em torno dos dois pontos de equilíbrio estável  $p$  e  $q$ ; a perturbação aleatória permite a passagem de uma bacia de atração para outra. Como a perturbação é "pequena", em geral esta passagem se deve à ocorrência de uma oscilação extraordinária — bem maior do que o usual — uma "grande flutuação". Isto determina seu caráter imprevisível. Como

$a(r) - a(p) < a(r) - a(q)$  tudo indica que o tempo passado próximo ao ponto  $p$  deve ser muito menor do que o passado próximo ao ponto  $q$ . Isto está de acordo com o fato de que se  $b(\cdot)$  crescer suficientemente no infinito, a única medida de probabilidade invariante  $\mu_\epsilon(\cdot)$ , dada por

$$\mu_\epsilon(dx) = c_\epsilon \exp(-\epsilon^{-2} 2a(x))dx, \quad \text{com } c_\epsilon^{-1} = \int_{\mathbb{R}^d} e^{-\frac{2a(x)}{\epsilon^2}} dx$$

e concentra-se cada vez mais em pequenas vizinhanças de  $q$ , quando  $\epsilon \downarrow 0$ .

Os Teoremas 3.2 e 3.3 darão enunciados rigorosos para o que dissemos, e aparecem em A. Galves et al. (1984). Aqui, nos preocuparemos mais com as idéias básicas, e em delinear a prova. Para demonstrações completas nos referimos ao artigo acima mencionado.

Há duas coisas a provar:

- (a) Exponencialidade do instante de transição.
- (b) Estabilidade das médias temporais.

Antes de enunciarmos os principais resultados desta seção seria conveniente fixarmos algumas notações: (1) Se  $C$  for um conjunto boreliano de  $\mathbb{R}^d$ ,  $\tau_\epsilon^x(C)$  denotará o instante da "primeira visita" a  $C$  pelo processo  $X_\epsilon^x(\cdot)$ , mais precisamente:

$$\begin{aligned} \tau_\epsilon^x(C) &= \inf\{t \geq 0: X_\epsilon^x(t) \in C\} \\ &= +\infty \quad \text{se tal conjunto for vazio.} \end{aligned}$$

(2)  $(\mathcal{G}_t)_{t \geq 0}$  denota a filtração natural no espaço  $(\Omega, \mathcal{G})$  associada ao movimento Browniano dado, ou seja,  $\mathcal{G}_t$  é a  $\sigma$ -álgebra gerada por  $(W_s: s \leq t)$  para cada  $t > 0$ . (Desta forma se  $C$  é aberto ou fechado,  $\tau_\epsilon^x(C)$  é um tempo de parada com relação a  $(\mathcal{G}_t)_{t \geq 0}$ , ou seja, para cada  $t > 0$  o evento  $[\tau_\epsilon^x(C) \leq t]$  pertence à  $\sigma$ -álgebra  $\mathcal{G}_t$ ).

(3) Como antes  $C_b(\mathbb{R}^d)$  denotará o espaço das funções reais contínuas e limitadas em  $\mathbb{R}^d$ , com a norma do supremo.

(4)  $m_1 (= m_1(\mathbb{R}^d))$  denotará o espaço das medidas de probabilidade com a topologia  $w^*$ .

Definição 3.1. Fixemos um  $c > 0$  tal que a bola de centro  $y$  e raio  $c$ ,  $\bar{B}_c(y)$ , esteja contida na bacia de atração de  $y$ , para  $y = p$  ou  $y = q$  e definamos os tempos de parada

$$T_\epsilon^x = \tau_\epsilon^x(B_c(q)) \quad (3.5)$$

para cada  $\epsilon > 0$  e  $x \in \mathbb{R}^d$ .

Podemos então enunciar:

Teorema 3.2. Para cada  $\epsilon > 0$ , seja  $\beta_\epsilon$  definido através da relação  $P_p(T_\epsilon > \beta_\epsilon) = e^{-1}$ . Então  $P_x(T_\epsilon/\beta_\epsilon > t) \xrightarrow[\epsilon \downarrow 0]{} e^{-t}$ , para todo  $t > 0$ , desde que  $x$  pertença à bacia de atração de  $p$ .

Teorema 3.3. Seja  $(\beta_\epsilon)_{\epsilon > 0}$  como no Teorema 3.2. Podemos encontrar  $R_\epsilon \uparrow +\infty$  de modo que os processos  $(v_t^\epsilon)_{t \geq 0}$ , a valores

em  $m_1(\mathbb{R}^d)$ , e definidos por

$$v_t^\epsilon(f) = \frac{1}{R_\epsilon} \int_{t\beta_\epsilon}^{t\beta_\epsilon + R_\epsilon} f(X_\epsilon^p(s)) ds, \quad f \in C_b(\mathbb{R}^d), \quad (3.4)$$

converjam em lei, quando  $\epsilon \downarrow 0$ , a um processo Markoviano de salto  $(v_t)_{t \geq 0}$  dado por:

$$\begin{aligned} v_t &= \delta_p \quad \text{se } t < T \\ &= \delta_q \quad \text{se } t \geq T, \end{aligned}$$

onde  $T$  é uma variável aleatória exponencial de média 1. De fato, podemos encontrar  $R_\epsilon \nearrow +\infty$  com  $R_\epsilon / \beta_\epsilon \rightarrow 0$  tais que:

$$P_p \left( \sup_{t < (T_\epsilon - 2R_\epsilon) / \beta_\epsilon} |v_t^\epsilon(f) - f(p)| > \delta \right) \rightarrow 0 \quad (3.5a)$$

$$P_p \left( \sup_{T_\epsilon / \beta_\epsilon < t < (\tilde{T}_\epsilon - 2R_\epsilon) / \beta_\epsilon} |v_t^\epsilon(f) - f(q)| > \delta \right) \rightarrow 0 \quad (3.5b)$$

para todo  $\delta > 0$  e toda  $f \in C_b(\mathbb{R}^d)$ , onde

$$\tilde{T}_\epsilon^x = \inf\{t > T_\epsilon : X_\epsilon^x(t) \in B_C(p)\}.$$

(Obs. O mesmo resultado vale se o ponto inicial pertence à bacia de atração de  $p$ .)

Antes de esboçarmos as provas dos teoremas que acabamos de enunciar, façamos algumas construções simples, que serão utili

zadas no decorrer de ambas as provas.

Devido à forma do potencial  $a(\cdot)$  (condições (i) - (iv)), para respondermos as questões acima delineadas é suficiente analisar o que acontece em um domínio limitado  $G$  de  $\mathbb{R}^d$ , que satisfaça as seguintes condições:

- (a)  $G$  tem bordo  $\partial G$  de classe  $C^2$ ;
- (b)  $B_c(p), B_c(q)$  estão contidos em  $G$ ;  $r \in G$ ;
- (c)  $\langle b(x), n(x) \rangle < 0$  para todo  $x \in \partial G$ , onde  $n(x)$  denota o vetor unitário normal e exterior a  $G$  no ponto  $x$ , e  $\langle \cdot, \cdot \rangle$  denota o produto interno euclidiano em  $\mathbb{R}^d$ . Em outras palavras,  $b(x)$  é transversal a  $\partial G$  em  $x$ , e aponta para o interior de  $G$ .

Pelas condições (b), (c) e a forma do potencial temos que  $a(r) < \min_{y \in \partial G} a(y)$ . Ademais, as hipóteses feitas implicam que  $W_r^S \cap G$  é conexo e divide  $G$  em dois subdomínios  $G \cap D_p$  e  $G \cap D_q$ . (Fig. 3.2).

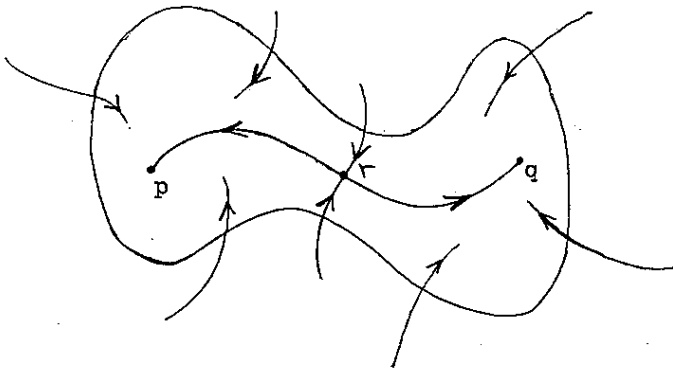


Fig. 3.2

Estamos interessados na passagem de  $G \cap D_p$  para  $G \cap D_q$ , pelo sistema perturbado. Note que no Capítulo 2 vimos resultados sobre a saída de um domínio cuja fronteira não continha pontos críticos. Este não é o caso de  $G \cap D_p$  e  $G \cap D_q$ . Por isto decomponemos  $G$  em subdomínios  $G_p, G_q$  e  $G_r$ , isolando pois uma "pequena" vizinhança de  $r$ . Para isto façamos uma observação bem simples, mas que nos será conveniente:

Observação. Dado  $\alpha > 0$  podemos encontrar  $\eta > 0$  tal que sendo

$$C \stackrel{\text{def}}{=} \{y: \text{distância}(y, W_r^S \cap G) < \eta\} \quad \text{temos}$$

$$\sup_{x \in C} E_x \tau_\epsilon(\mathbb{R}^d \setminus C) < e^{\alpha/\epsilon^2} \quad (3.6a)$$

para  $\epsilon$  suficientemente pequeno.

Prova da observação. É bem simples, e análoga ao Lema 1.7 do Capítulo 6 de [F-W]. Antes de tudo pela condição de Lipschitz do campo  $b(\cdot)$  sabemos que existe  $L < +\infty$  tal que para todo  $x, y \in \mathbb{R}^d$  podemos encontrar  $\varphi(\cdot)$  com  $\varphi(0) = x$ ,  $\varphi(T) = y$ , onde  $T = |y-x|$  e  $S_T(\varphi) \leq L|x-y|$ , pelo Lema 2.9.

Seja  $0 < \tilde{\eta} < \alpha/6L$  e seja  $\tilde{T}_0 < +\infty$  tal que se  $x \in W_r^S \cap G$  então  $X^x(\cdot)$  alcança  $B_{\tilde{\eta}}(r)$  antes do tempo  $\tilde{T}_0$ . Fixemos  $z$  com  $|z-r| = \text{distância}(z, W_r^S \cap G) = \tilde{\eta}$  e seja  $\eta = \tilde{\eta}/2$ .

Se  $C \stackrel{\text{def}}{=} \{x: \text{distância}(x, W_r^S \cap G) < \eta\}$  podemos encontrar, para cada  $x \in C$ ,  $\varphi_x(\cdot)$  tal que  $\varphi_x(0) = x$  e  $\varphi_x(T) = z$  para certo  $T \leq T_0 = \tilde{T}_0 + 4\eta$  e com  $S_T(\varphi_x) \leq 3L\tilde{\eta} < \alpha/2$  e então  $P_{T_0}(X_\epsilon(\cdot), \varphi_x) < \eta$  implica  $\tau_\epsilon^x(\mathbb{R}^d \setminus C) < T_0$ . Daí temos a existência de

$\epsilon_0 > 0$  de modo que se  $\epsilon < \epsilon_0$

$$\begin{aligned} \inf_{x \in C} P_x(\tau_\epsilon(\mathbb{R}^d \setminus C) < T_0) &\geq \inf_{x \in C} P_x(p_{T_0}(X_\epsilon(\cdot), \varphi_x) < \delta) \geq \\ &\geq e^{-\frac{3\alpha}{4\epsilon^2}} \end{aligned}$$

(Exerc.: Justifique isto.)

Utilizando a propriedade de Markov teremos para  $\epsilon < \epsilon_0$

$$\sup_{x \in C} P_x(\tau_\epsilon(\mathbb{R}^d \setminus C) > nT_0) \leq \left(1 - e^{-\frac{3\alpha}{4\epsilon^2}}\right)^n$$

de modo que

$$\begin{aligned} \sup_{x \in C} E_x(\tau_\epsilon(\mathbb{R}^d \setminus C)) &\leq T_0 e^{\frac{3\alpha}{4\epsilon^2}} \leq \\ &\leq e^{\alpha/\epsilon^2} \end{aligned}$$

se  $\epsilon$  for suficientemente pequeno e isto conclui a prova da observação.

Comentário. Esta simples observação não usa totalmente o fato de  $r$  ser uma sela; aqui podemos ter  $\eta > 0$  dependendo do  $\alpha > 0$  fixado. Mais tarde, para o Teorema 3.3 precisaremos de algo mais forte, o lema 5(a) de [G.O.V.]. Por enquanto faremos uma escolha preliminar fixando um certo  $0 < \alpha_0 < 2(a(r) - a(p))$  e tomando  $\eta_0 > 0$  correspondente, de acordo com a observação.

Feita tal escolha de  $\eta_0 > 0$  faremos a decomposição de  $G$  tomando variedades  $C^2$ ,  $(d-1)$ -dimensionais,  $\Gamma$  e  $\tilde{\Gamma}$ , de modo que:  $\Gamma \subseteq G \cap D_p$ ,  $\Gamma$  separa  $G$  em apenas dois subdomínios

$G_p$  e  $\tilde{\Lambda}$  sendo que  $\bar{B}_c(p) \subset G_p \subset G \cap D_p$ ; analogamente  $\tilde{\Gamma} \subset G \cap D_q$  e  $\tilde{\Gamma}$  separa  $G$  em apenas dois subdomínios  $G_q$  e  $\Lambda$ , sendo que  $\bar{B}_c(q) \subset G_q \subset G \cap D_q$ . Exigimos ainda que para cada  $x \in \Gamma \cup \tilde{\Gamma}$  sua distância a  $W_r^S \cap G$  seja inferior a  $\eta_0$ , acima fixado. Definimos  $G_r = \Lambda \cap \tilde{\Lambda}$ .

Para melhor esclarecer cabem algumas observações:

- (a) Nestas condições o bordo  $\partial G_r$  contém  $\Gamma$  e  $\tilde{\Gamma}$  e duas componentes conexas do bordo  $\partial G$ .
- (b)  $G$  é união disjunta de  $G_p, G_q, G_r, \Gamma$  e  $\tilde{\Gamma}$ .
- (c) As variedades  $\Gamma$  e  $\tilde{\Gamma}$  podem ser pensadas como transladas de  $W_r^S \cap G$ , ou como pedaços de hiperplanos convenientes.
- (d) Sendo  $\Gamma$  e  $\tilde{\Gamma}$  relativamente compactos, podemos tomar  $M < +\infty$  tal que se  $x \in \Gamma(\tilde{\Gamma})$ , a trajetória determinística  $X^x(\cdot)$  alcança  $B_{c/2}(p)$  ( $B_{c/2}(q)$ , respectivamente) em um tempo inferior a  $M$ . (Ver Fig.3.3).
- (e) Pela escolha de  $\eta_0$  e por (3.6) teremos

$$\sup_{x \in G_r} E_x \tau_\epsilon(G_r^c) < e^{\alpha_0/\epsilon^2} \quad (3.6b)$$

onde  $\alpha_0 < 2(a(r) - a(p))$ , o que dá um tempo (médio) infinitamente menor, quando  $\epsilon \downarrow 0$ , do que  $E_p(T_\epsilon)$ .



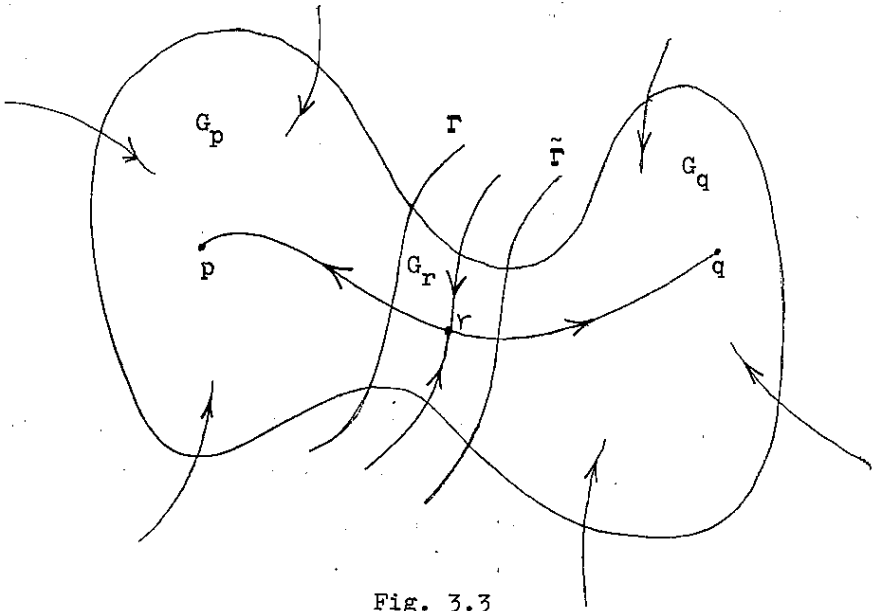


Fig. 3.3

Prova do Teorema 3.2: (O tempo exponencial) Observe inicialmente que para  $x \in \Lambda$  o tempo necessário para sair de  $\Lambda$  não difere muito de  $T_\epsilon$ ; com efeito se  $S_\epsilon^x = \tau_\epsilon^x(\partial\Lambda)$ ,  $x \in \Lambda$ , então  $S_\epsilon^x \leq T_\epsilon^x$  e

$$P_x(T_\epsilon \leq S_\epsilon + M) \xrightarrow[\epsilon \downarrow 0]{} 0 \quad (3.7)$$

De fato, pela Teoria de Freidlin e Wentzell sabemos que  $P_x(X_\epsilon(S_\epsilon) \in \tilde{\Gamma}) \rightarrow 1$  quando  $\epsilon \downarrow 0$ ; agora, se em  $\tilde{\Gamma}$ , a probabilidade de que  $X_\epsilon(\cdot)$  alcance  $B_c(q)$  em tempo não superior a  $M$  tende a um, já que neste tempo o sistema determinístico alcança  $B_{c/2}(q)$  e porque vale (2.8). Combinando isto, através da propriedade de Markov, (3.7) decorre.

Esta afirmação nos permite substituir o Teorema 3.2 pela Proposição abaixo. (Detalhes em Galves, Olivieri, Vares (1984)\*.)

Proposição 3.4. Definamos  $\gamma_\epsilon$  através de  $P_p(S_\epsilon > \gamma_\epsilon) = e^{-1}$ . Então  $P_x(S_\epsilon > t\gamma_\epsilon) \rightarrow e^{-t}$  para todo  $x \in G_p$ .

Prova: (idéia básica; detalhes em [GOV].)

Começemos com  $x = p$ . Gostaríamos de verificar que

$$P_p(S_\epsilon/\gamma_\epsilon > t+s) - P_p(S_\epsilon/\gamma_\epsilon > t) P_p(S_\epsilon/\gamma_\epsilon > s) \rightarrow 0$$

para todo  $s, t$ , quando  $\epsilon \rightarrow 0$ . (Isto implicaria o resultado no caso  $x = p$ ). Por outro lado, uma coisa natural é calcular

$P_p(S_\epsilon > (t+s)\gamma_\epsilon)$  condicionando na posição do processo no instante  $t\gamma_\epsilon$ , que é um ponto em  $\Lambda$ :

$$P_p(S_\epsilon > t\gamma_\epsilon + s\gamma_\epsilon) = \int_{\Lambda} P_y(S_\epsilon > s\gamma_\epsilon) P_p(S_\epsilon > t\gamma_\epsilon, X_\epsilon(t\gamma_\epsilon) \in dy)$$

Se  $P_y(S_\epsilon > t\gamma_\epsilon) \sim P_p(S_\epsilon > t\gamma_\epsilon)$  para  $y \in \Lambda$  estaria tudo pronto. Entretanto,  $\Lambda$  contém pontos do domínio de atração de  $q$  para os quais  $S_\epsilon$  é muito menor, de modo que o argumento não é válido. No entanto precisamos (e provamos) apenas algo mais fraco, contido em (i) e (ii) abaixo enunciados:

(i) Existem  $\eta_\epsilon > 0$ ,  $\eta_\epsilon/\gamma_\epsilon \rightarrow 0$ , tais que

$$\sup_{y \in B_c(p)} P_y(S_\epsilon > t\gamma_\epsilon) \leq P_p(S_\epsilon > t\gamma_\epsilon - \eta_\epsilon) + o(1) \quad (3.8a)$$

\* Esta referência será indicada por [G.O.V.]

$$\inf_{y \in \bar{B}_C(p)} P_y(S_\epsilon > t\gamma_\epsilon) \geq P_p(S_\epsilon > t\gamma_\epsilon + \eta_\epsilon) - o(1) \quad (3.8b)$$

onde  $o(1) \rightarrow 0$  quando  $\epsilon \rightarrow 0$  e podemos tomá-la independente de  $(t > 0)$ .

(ii) No caso  $X_\epsilon(t\gamma_\epsilon) \in \Lambda \setminus \bar{B}_C(p)$  mas  $S_\epsilon > t\gamma_\epsilon + s\gamma_\epsilon$ , tomemos  $\eta'_\epsilon = \exp(\alpha'/\epsilon^2)$  onde  $\alpha_0 < \alpha' < 2(a(r) - a(p))$ , de modo que  $\eta'_\epsilon / \gamma_\epsilon \rightarrow 0$ , e examinemos as duas possibilidades existentes se  $s\gamma_\epsilon > \eta'_\epsilon$ ; ou  $X_\epsilon(\cdot)$  permanece em  $\Lambda \setminus \bar{B}_C(p)$  durante  $[t\gamma_\epsilon, t\gamma_\epsilon + \eta'_\epsilon]$  ou volta a  $\bar{B}_C(p)$  antes de  $t\gamma_\epsilon + \eta'_\epsilon$ . A primeira situação tem probabilidade que converge a zero, como veremos utilizando (3.6b).

Na segunda possibilidade condicionamos nos possíveis valores de  $(R^t, X_\epsilon(R^t))$ , onde  $R^t = \inf\{u > t\gamma_\epsilon : X_\epsilon(u) \in \bar{B}_C(p)\}$  e utilizamos a propriedade forte de Markov, de modo a obtermos: ( $o_s(1) \rightarrow 0$  se  $\epsilon \rightarrow 0$ , para cada  $s > 0$ )

$$(3.9a)$$

$$P_p(S_\epsilon > (t+s)\gamma_\epsilon) \leq P_p(S_\epsilon > t\gamma_\epsilon) \sup_{y \in \bar{B}_C(p)} P_y(S_\epsilon > s\gamma_\epsilon - \eta'_\epsilon) + o_s(1)$$

$$(3.9b)$$

$$P_p(S_\epsilon > (t+s)\gamma_\epsilon) \geq P_p(S_\epsilon > t\gamma_\epsilon + \eta'_\epsilon) \inf_{y \in \bar{B}_C(p)} P_y(S_\epsilon > s\gamma_\epsilon) - o_s(1)$$

quando  $\epsilon \rightarrow 0$ , de modo que (3.8) possa ser utilizada.

Verificação de (i) e (ii):

Para (3.8) tomamos  $K$  um compacto tal que  $\bar{B}_C(p) \subset \overset{\circ}{K} \subset K \subset G_p$ ; M. Day (1983) provou que

$$\sup_{y \in \bar{B}_C(p)} \|\mu_y^\epsilon - \mu_p^\epsilon\| \xrightarrow{\epsilon \downarrow 0} 0$$

onde  $\|\cdot\|$  é a norma da variação total e  $\mu_y^\epsilon$  é a lei de  $X_\epsilon^y(\tau_\epsilon(\partial K))$ , i.e., a distribuição com que  $X_\epsilon^y(\cdot)$  atinge  $\partial K$  pela primeira vez ( $y \in \bar{B}_C(p)$ ). Este fato permite (cf. [G.O.V.]) a construção de um acoplamento de  $X_\epsilon^y$  e  $X_\epsilon^p$  de modo que, com probabilidade tendendo a um ( $\epsilon \downarrow 0$ ), eles atingem  $\partial K$  no mesmo ponto (em tempos possivelmente diferentes) e então percorrem a mesma trajetória, de modo que a diferença entre seus tempos de chegada a  $\partial \Lambda$  provém apenas da parte inicial, i.e., até chegar em  $\partial K$ . Entretanto, sendo

$$\inf_{y \in \partial K} |a(y) - a(p)| < a(r) - a(p)$$

então  $\tau_\epsilon(\partial K) \ll S_\epsilon$  de acordo com (2.40), (2.41), i.e. devem diferir no máximo por  $\eta_\epsilon \ll \gamma_\epsilon$ , em um conjunto com probabilidade tendendo a um. Isto dá as desigualdades (3.8a) e (3.8b).

Para concluirmos a prova precisamos verificar (3.9a) e (3.9b). Para isto seja  $\eta'_\epsilon$  como em (ii) acima. Temos então:

$$\begin{aligned} & \sup_{y \in \Lambda \setminus \bar{B}_C(p)} P_y(S_\epsilon > s\gamma_\epsilon, \tau_\epsilon(\bar{B}_C(p)) > \eta'_\epsilon) \\ & \leq \sup_{y \in \bar{G}_p} P_y(\tau_\epsilon(\bar{B}_C(p)) > \eta'_\epsilon) + \sup_{y \in \bar{G}_r} P_y(S_\epsilon > s\gamma_\epsilon, \tau_\epsilon(\bar{B}_C(p)) > \eta'_\epsilon) \end{aligned}$$

Se  $M_0 < +\infty$  for tal que para cada  $x \in \bar{G}_p$ ,  $X^x(\cdot)$  atinge  $B_{C/2}(p)$  antes do instante  $M_0$ , por (2.8) teremos:

$$\sup_{y \in \bar{G}_p} P_y(\tau_\epsilon(B_C(p)) > M_0) \rightarrow 0 \quad (*)$$

o que em particular mostra que o primeiro termo da desigualdade tende a zero. Quanto ao segundo, se  $s\gamma_\epsilon > \eta'_\epsilon/2$  podemos decompor o evento  $[S_\epsilon > s\gamma_\epsilon, \tau_\epsilon(\bar{B}_C(p)) > \eta'_\epsilon]$  de acordo com as duas possibilidades: ou  $X^y_\epsilon(\cdot)$  permanece em  $\bar{G}_r$  durante  $[0, \eta'_\epsilon/2]$  ou alcança  $\bar{G}_p$  em  $[0, \eta'_\epsilon/2]$ . Obtemos então pela propriedade forte de Markov, se  $s\gamma_\epsilon > \eta'_\epsilon/2$

$$\begin{aligned} & \sup_{y \in \bar{G}_r} P_y(S_\epsilon > s\gamma_\epsilon, \tau_\epsilon(B_C(p)) > \eta'_\epsilon) \leq \\ & \leq \sup_{y \in \bar{G}_r} P_y(\tau_\epsilon(G_r^c) > \eta'_\epsilon/2) + \sup_{y \in \bar{G}_p} P_y(\tau_\epsilon(B_C(p)) > \eta'_\epsilon/2), \end{aligned}$$

e ambos tendem a zero, por (3.6) e por (\*) acima. Com isto, se definimos  $R^t = \inf\{u > t\gamma_\epsilon : X_\epsilon(u) \in \bar{B}_C(p)\}$ , pela propriedade de Markov teremos:

$$\sup_{y \in \Lambda} P_y(S_\epsilon > t\gamma_\epsilon + s\gamma_\epsilon, R^t > t\gamma_\epsilon + \eta'_\epsilon) \rightarrow 0.$$

Então, condicionando nos possíveis valores de  $(R^t, X_\epsilon(R^t))$  e usando novamente a propriedade forte de Markov obtemos (3.9a) e (3.9b). ■

Estabilidade das médias temporais (esboço da prova)

Sejam  $G_p, G_q, G_r, \Gamma, \tilde{\Gamma}, \Lambda, \tilde{\Lambda}$  como antes. Para a prova a seguir, o ingrediente básico mais importante é uma estimativa da proporção de tempo passada fora de  $\bar{B}_\theta(p)$ , onde  $\theta > 0$  é arbitrariamente pequeno, enquanto o processo ainda não saiu de  $\Lambda$ :

Lema 3.5. Sejam  $0 < \alpha < 2(a(r)-a(p))$  e  $R'_\epsilon = \exp(\alpha/\epsilon^2)$ . Então, para todo  $\theta > 0, \delta > 0$  existe  $\epsilon_0 = \epsilon_0(\theta, \delta) > 0$  tal que: ( $I_\theta(\cdot)$  denota a função indicadora de  $\bar{B}_\theta(p)$ ).

$$\sup_{x \in \Lambda} P_x \left[ \left| \frac{1}{R} \int_0^R I_\theta(X_\epsilon(s)) ds - 1 \right| > \delta, S_\epsilon > R \right] \leq \exp(-c_\delta R/R'_\epsilon) \tag{3.10}$$

se  $R > 2R'_\epsilon$  e  $\epsilon \leq \epsilon_0$ , onde  $c_\delta$  é constante positiva (dependendo de  $\delta$ ).

Prova: (cf. [G.O.V.] (1984))

Dados  $\theta, \delta$  e  $\alpha$  como acima, tomemos  $\theta_0$  tal que  $0 < \theta_0 \leq \theta$  e  $0 < \min_{z: |z-p|=\theta_0} (a(z)-a(p)) \leq \alpha$ , e seja

$t_\epsilon = \exp(\epsilon^{-2} \min_{z: |z-p|=\theta_0} (a(z) - a(p)))$ , de modo que  $t_\epsilon \leq R'_\epsilon$ . Utilizaremos então a decomposição:  $(0, R] = \bigcup_{i=1}^{N_\epsilon} ((i-1)t_\epsilon, it_\epsilon] \cup (N_\epsilon t_\epsilon, R]$

onde  $N_\epsilon = [R/t_\epsilon]$ . Se  $t_\epsilon/R < \delta/2$ , então

$$\left[ \left| \frac{1}{R} \int_0^R I_{\theta_0}(X_{\epsilon}(s)) ds - 1 \right| > \delta, S_{\epsilon} > R \right] \subset$$

$$\subset \left[ \left| \frac{1}{N_{\epsilon} t_{\epsilon}} \int_0^{N_{\epsilon} t_{\epsilon}} I_{\theta_0}(X_{\epsilon}(s)) ds - 1 \right| > \delta/2, S_{\epsilon} > N_{\epsilon} t_{\epsilon} \right] \quad (3.11)$$

Podemos tomar  $K_1, K_2$  compactos (com  $\partial K_i$  de classe  $C^2$ ) tais que

$$(i) \quad p \in \overset{\circ}{K}_1 \subset K_1 \subset \overset{\circ}{K}_2 \subset B_{\theta_0}(p)$$

$$(ii) \quad 2 \min_{z \in \partial K_2} (a(z) - a(p)) > \min_{|z-p|=\theta_0} (a(z) - a(p))$$

(iii)  $\langle b_0(x), n_i(x) \rangle < 0$  para  $x \in \partial K_i$ , onde  $n_i(x)$  é o vetor normal (unitário) exterior a  $\partial K_i$  em  $x$ ,  $i=1,2$ .

Desta forma, se definirmos

$$p(\epsilon) \stackrel{\text{def}}{=} \sup_{x \in K_1} P_x(\tau_{\epsilon}(K_2) < t_{\epsilon}) + \sup_{x \in \Lambda} P_x(\tau_{\epsilon}(K_1) > \sqrt{t_{\epsilon}}, S_{\epsilon} > t_{\epsilon})$$

então  $p(\epsilon) \rightarrow 0$  quando  $\epsilon \rightarrow 0$ , pelos resultados de Freidlin e Wentzell que já discutimos. Ademais, se definimos

$$Y_{\epsilon}^i(w) = 0 \quad \text{se } X_{\epsilon}^p(\cdot) \text{ visita } K_1 \text{ em } ((i-1)t_{\epsilon}, (i-1)t_{\epsilon} + \sqrt{t_{\epsilon}}] \\ \text{e depois passa o restante do intervalo } \\ ((i-1)t_{\epsilon}, it_{\epsilon}] \text{ em } K_2$$

$$Y_{\epsilon}^i(w) = 1, \quad \text{caso contrário;}$$

então

$$P(Y_{\epsilon}^{i_1} = 1, \dots, Y_{\epsilon}^{i_k} = 1, S_{\epsilon} > N_{\epsilon} t_{\epsilon}) \leq p(\epsilon)^k \quad (3.12)$$

para todo  $k \geq 1$  e todos  $i_1, \dots, i_k \leq N_{\epsilon}$  distintos. Por outro lado

$$\left[ \sum_{i=1}^{N_{\epsilon}} Y_{\epsilon}^i \leq k \right] \leq \left[ \left| \frac{1}{N_{\epsilon} t_{\epsilon}} \int_0^{N_{\epsilon} t_{\epsilon}} I_{\theta_0}(X_{\epsilon}(s)) ds - 1 \right| \leq \frac{kt_{\epsilon} + (N_{\epsilon} - k)\sqrt{t_{\epsilon}}}{N_{\epsilon} t_{\epsilon}} \right]$$

Portanto de (3.11) e (3.12); se  $\sqrt{t_{\epsilon}} > 4/\delta$  e  $t_{\epsilon}/R < \delta/2$ :

$$P\left[ \left| \frac{1}{R} \int_0^R I_{\theta}(X_{\epsilon}(s)) ds - 1 \right| > \delta, S_{\epsilon} > R \right] \quad (3.13)$$

$$\leq P\left[ \sum_{i=1}^{N_{\epsilon}} Y_{\epsilon}^i > \frac{\delta}{4} N_{\epsilon}, S_{\epsilon} > N_{\epsilon} t_{\epsilon} \right]$$

$$\leq \sum_{k > \frac{\delta}{4} N_{\epsilon}} \binom{N_{\epsilon}}{k} p(\epsilon)^k.$$

Isto lembra a probabilidade de uma binomial com parâmetros  $N_{\epsilon}$  e  $p(\epsilon)$  ser maior do que  $\frac{\delta}{4} N_{\epsilon}$ , o que sabemos tender a zero exponencialmente (desigualdade de Chebychev exponencial, cf. §1.1). Entretanto não é bem isto que temos em (3.13): está faltando o fator  $(1-p(\epsilon))^{N_{\epsilon}-k}$ , em cada parcela no lado direito de (3.13). Mas como estamos interessados apenas na cauda da distribuição é possível ainda efetuar a comparação com binomial de parâmetros  $N_{\epsilon}$  e  $p'(\epsilon)$  onde ainda  $p'(\epsilon) \rightarrow 0$ , bastando que  $p(\epsilon) \leq p'(\epsilon)(1-p'(\epsilon))^{4/\delta}$ . Obtem-se então



$$P \left[ \left| \frac{1}{R} \int_0^R I_{\theta}(X_{\epsilon}(s)) ds - 1 \right| > \delta, S_{\epsilon} > R \right] \leq \exp \left( -\frac{\delta}{8} N_{\epsilon} \right) \\ \leq e^{-\frac{\delta}{16} \frac{R}{R_{\epsilon}'}}$$

se  $\epsilon$  for suficientemente pequeno e  $R/R_{\epsilon}' \geq 2$ .

Prova do Teorema 3.3. Uma vez provadas (3.5a) e (3.5b) o Teorema decorre por argumentos gerais [G.O.V. (1984)], utilizando-se também o Teorema 3.2. Na prova de (3.5a) e (3.5b) a estimativa importante vem do Lema 3.5 (utilizado duas vezes: com relação a  $p$  e depois com relação a  $q$ ). Para os detalhes nos referimos ao artigo acima mencionado ( $R_{\epsilon}$  pode ser tomado como  $\beta_{\epsilon}^{\gamma}$  se  $0 < \gamma < 1$ ). Note que como  $\tilde{T}_{\epsilon}/\beta_{\epsilon} \rightarrow +\infty$  em probabilidade, na escala temporal em que tomamos  $v_t^{\epsilon}$ , não poderemos ver o retorno a  $B_c(p)$  após visitar  $B_c'(q)$ .

Comentários adicionais. Em Cassandro et. al (1984), onde foi apresentada esta proposta com base nas médias temporais de trajetórias típicas, há a discussão de um outro modelo apresentando comportamento metaestável. É o chamado processo de Contacto (básico) de Harris,  $(\xi(t))_{t \geq 0}$  tomando valores em  $\{0,1\}^{\mathbb{Z}}$ , com a interpretação original sendo de que as posições  $x \in \mathbb{Z}$  representam indivíduos, e  $\xi(t,x) = 1$  ou  $0$  significando que o indivíduo em  $x$  está infectado ou não respectivamente. A dinâmica depende de um parâmetro  $\lambda > 0$  que representa a taxa com que um

indivíduo infectado pode infectar cada um de seus dois vizinhos. Ademais a recuperação de um indivíduo infectado ocorre com taxa 1 (tempo exponencial de média 1). Para a construção e propriedades básicas deste processo referimos a Griffeath (1979) ou à rezenha Griffeath (1981). Ocorre que existe  $0 < \lambda_c < +\infty$  t.q. se  $\lambda > \lambda_c$  o processo possui duas medidas de probabilidade invariantes extremas (i.e., as outras são combinações convexas), que são  $\delta_\emptyset$  i.e., medida de Dirac na configuração  $\xi(x) \equiv 0$ , e uma outra não trivial  $\nu_\lambda$ , que pode ser obtida como limite em lei de  $\xi(t)$  quando  $\xi(0) = Z$  e quando  $t \rightarrow +\infty$ . Quando truncamos o processo com uma redução aos indivíduos  $\{-N, \dots, N\}$  passamos a ter apenas uma situação de equilíbrio: todos são (i.e.  $\delta_\emptyset$ ). Entretanto se  $N$  é grande,  $\lambda > \lambda_c$ , e começamos com  $\xi_N(0) = \{-N, \dots, N\}$ , a maneira como este equilíbrio é atingido ocorre de forma análoga à passagem de  $B_c(p)$  para  $B_c(q)$  no exemplo anterior; tem-se a impressão de que  $\nu_\lambda$  restrita a  $\{0, 1\}^{\{-N, \dots, N\}}$  seja uma medida invariante pois médias temporais convenientes estabilizam-se em torno desta medida; entretanto após um tempo muito longo e de forma imprevisível (i.e. tempo aproximadamente exponencial) ocorrerá a transição para a configuração vazia. O método como obtemos as estimativas básicas neste exemplo é bem diferente; entretanto recomendamos que o leitor aí detecte o papel de estimativas de grandes desvios, que continuam presentes. Este exemplo apresenta uma estrutura espacial que permite interessantes questões sobre o intervalo (nos quais estamos tomando as médias temporais) que contém o instante da absorção pela configuração vazia, e.g., há formação de "gota crítica"

ou não? Tudo sugere a segunda opção, ao contrário do que se espera em modelos (mais interessantes para Mecânica Estatística) tais como o Modelo de Ising bidimensional para o qual ainda não se conseguiu um resultado análogo ao Teorema 3.3.

Extensões dos resultados sobre o processo de contacto contidos em Cassandro et. al (1984) foram obtidos por Roberto Schonmann (1985).

### §3.3 Comportamento metaestável em sistemas com dinâmica determinística.

Após a análise do modelo  $X_e(\cdot)$  da secção 3.2 é natural questionar-se sobre a possibilidade de se observar comportamento semelhante em sistemas com dinâmica determinística, i.e., sem a introdução de uma perturbação aleatória (ou ruído aleatório externo).

Um grupo de pesquisadores, incluindo A. Lasota, G. Pianigiani, J.A. Yorke e E. Yorke<sup>(1)</sup> têm desenvolvido uma série de trabalhos onde consideram situações deste tipo, a que chamam de "caos metastável" ou "comportamento caótico transitório". Nessas situações pode ocorrer que o sistema se comporte por um tempo muito longo como se fosse descrito por uma medida invariante  $\mu$ , absolutamente contínua em relação à medida de Lebesgue. Entretanto "subitamente" ocorrerá uma brusca mudança no seu comportamento podendo então ser rapidamente atraído a um conjunto de medida

---

(1) Para referências mais completas consulte [22] ou ainda [28].

de Lebesgue zero (e portanto nulo com relação à antiga  $\mu$ ). Há exemplos interessantes que ocorrem em conexão com modelos de dinâmica dos fluídos. Este comportamento ("caos metaestável") aparece em um período que pode ser chamado pré-turbulento (Eckmann, Ruelle (1985)). Mencionemos um exemplo concreto: o sistema de equações diferenciais de Lorenz:

$$\frac{dx}{dt} = \sigma y - \sigma x$$

$$\frac{dy}{dt} = -xz + rx - y$$

$$\frac{dz}{dt} = xy - bz$$

$$(\sigma = 10; \quad b = 8/3)$$

Aqui o comportamento é simples quando  $r$  é pequeno. Se  $1 < r < r_0$  ( $\approx 13,93$ ), há três pontos fixos, dois estáveis e um instável e todas as trajetórias convergem para um deles. Por outro lado cf. observado por Lorenz se  $R \approx 28$  há um comportamento que pode ser chamado de caótico ou turbulento (Lorenz (1963); ver também Mardsen McCracken 1976) para discussões bem mais detalhadas.) Foi observado por Robbins (1977) e Kaplan e Yorke (1977) que há um ponto  $r_1 \approx 24,06$  que representa de certa forma uma transição entre estes dois tipos mais extremos de comportamento. Se  $r < r_1$  mas  $r$  for muito próximo a  $r_1$  observa-se um comportamento aparentemente caótico por um longo tempo, mas finalmente a trajetória tende a um ponto fixo. É este um exemplo

de grande interesse; entretanto, como no caso do modelo de Ising bidimensional, quando nos referíamos a dinâmicas estocásticas os resultados já conhecidos sobre "comportamento metaestável" não se aplicarão a ele. Estaremos pois nos limitando a situações mais simples. Os resultados existentes ainda são de certo modo preliminares, e há muitas questões por resolver.

Uma outra classe de exemplos está ligada a algoritmos para gerar números "pseudo aleatórios em  $[0,1]$ , como o método do "quadrado médio" de Newmann. São sucessivas iterações de uma aplicação, onde pela observação das trajetórias podemos ficar con- vencidos de termos valores com distribuição aproximadamente uniforme em  $[0,1]$  e, no entanto, após número suficientemente grande de iterações as observações se concentram em subconjuntos bem pequenos do  $[0,1]$ . (Pianigiani, Yorke (1979).)

Em estudos numéricos para uma classe de exemplos J. Yorke e E. Yorke (1979) observaram que a distribuição do tempo que decorre até a mudança de comportamento deveria ser aproximado por uma variável aleatória exponencial com pequeno erro, para uma ampla classe de distribuições (aleatórias) iniciais. O aparecimento da lei exponencial fica facilmente entendido uma vez que se identifique a "transição" como decorrência de uma grande flutuação, uma sucessão de eventos improváveis, ou enfim algo que nos faz pensar em um cara-coroa onde coroa é muito mais provável e es- peramos pelo aparecimento da primeira cara. A tentativa de se provar um resultado deste gênero fica ligada naturalmente à verificação de perda de memória i.e., para  $t$  grande, dada a não-ocorrência da transição em  $(0,t]$  a situação no instante  $t$  no

que diz respeito ao tempo que falta para tal ocorrência é próxima à inicial. (É análogo à Proposição 3.4). Na verdade, ao provarem "decaimento exponencial" Lasota e Yorke utilizam algo até mais forte contido no conceito de medida "condicionalmente invariante".

Os resultados existentes referem-se a aplicações "expansoras"  $T: \bar{A} \rightarrow \mathbb{R}^d$  onde  $A \subset \mathbb{R}^d$  é um conjunto aberto, relativamente compacto, e com um número finito de componentes conexas. O sistema dinâmico em questão consiste em sucessivas iterações  $x \in A, Tx, T^2x, \dots$ ; é claro que só faz sentido iterar  $T$  enquanto a trajetória permanece em  $\bar{A}$ . Na situação sob consideração, se a distribuição inicial é absolutamente contínua em  $A$ , o tempo necessário para sair de  $A$  é grande e a dinâmica é tal que um certo "equilíbrio" se estabelece enquanto em  $A$ , embora  $A$  não seja invariante e finalmente as trajetórias saiam de  $A$ , terminando o processo. A pergunta é então: o que significa este "falso equilíbrio"? Estamos nos referindo, de modo um tanto informal, aos seguintes itens:

(a) Existência de medida de probabilidade  $\mu_0$  em  $A$  (absolutamente contínua) que seja condicionalmente invariante, i.e., se  $X$  é variável aleatória em  $A$  com distribuição  $\mu_0$ , a distribuição condicional de  $T(X)$  dado que  $T(X) \in A$  é também  $\mu_0$ .

(b) Se  $\tau(x) = \min\{n \geq 1: T^n x \notin A\}$  e a distribuição inicial  $\eta$  é absolutamente contínua,  $d\eta = fd\lambda$ , com  $f$  suave, então a lei de  $T^n X | \tau > n$  tende a  $\mu_0$  e obtém-se boa aproximação já para valores de  $n$  muito menores do que  $E\tau$ , i.e., com

$\eta(\tau > n)$  quase um.

A seguir tratamos de melhor definir o quadro descrito através de uma breve exposição dos resultados contidos em Piani-giani, Yorke (1979), sobre medidas condicionalmente invariante pa- ra transformações expansoras e sua aplicação ao decaimento expo- nencial (cf. Lasota, Yorke (1981)). Conjectura-se a validade de resultados análogos em situações bem mais gerais (Eckmann, Ruelle (1985), §3.5).

Definição 3.6. Seja  $T: \bar{A} \rightarrow \mathbb{R}^d$  uma transformação, onde

$$A = \bigcup_{i=1}^p A_i \subseteq \mathbb{R}^d$$

com  $A_1, \dots, A_p$  conjuntos abertos, conexos e dois a dois disjun- tos. Supõe-se que para cada  $A_i$  exista  $\delta_i < +\infty$  tal que cada dois pontos de  $A_i$  podem ser ligados por uma poligonal contida em  $A_i$  de comprimento não superior a  $\delta_i$ . Supõe-se ainda que  $T$  seja de classe  $C^2$  em  $A$  e que:

- (1)  $A \subseteq T(A)$
- (2) Se  $x_n \rightarrow x \in \partial A$  e  $Tx_n \rightarrow y$  então  $y \notin A$ .
- (3) Existe  $\alpha > 1$  tal que para todo  $x \in A$  a matriz Jaco- biana  $DT(x)$  é  $\alpha$ -expansora, onde dizemos que uma matriz  $M$  é  $\alpha$ -expansora se  $\inf_{|v|=1} |M \cdot v| \geq \alpha$ .

Uma aplicação  $T$  nas condições acima é chamada de aplicação expansora.

Definição 3.7. Seja  $T$  como na definição anterior. Uma medida de probabilidade  $\mu$  em  $A$  é dita ser condicionalmente invariante se  $\mu(T^{-1}(A)) > 0$  e

$$\mu(B) = \frac{\mu(T^{-1}(B))}{\mu(T^{-1}(A))}$$

para todo  $B$  boreliano, contido em  $A$ . (Ou seja se  $X$  é variável aleatória com distribuição  $\mu$  então a distribuição condicional de  $T(X)$  dado  $T(X) \in A$  é também  $\mu$ .)

Precisaremos ainda da seguinte definição.

Definição 3.8. Seja  $T$  como na Definição 3.7. (a) Dizemos que  $T$  é transitiva nas componentes se para cada par  $(i, j) \in \{1, \dots, p\}^2$  existe  $n = n(i, j) \geq 1$  tal que  $A_i \cap T^n(A_j) \neq \emptyset$ . (b) Se existir  $r \in \mathbb{N}$  tal que  $A \subset T^r(A_i)$  para todo  $i = 1, \dots, p$  dizemos que  $T$  é exata. (Obviamente, exata  $\rightarrow$  transitiva nas componentes.)

Algumas notações. (a) Como antes  $\lambda(\cdot)$  denotará a medida de Lebesgue em  $\mathbb{R}^d$  e vamos supor, sem perda de generalidade, que  $\lambda(A) = 1$ . (b)  $C_b(A) = \{f: A \rightarrow \mathbb{R} \mid f \text{ é contínua e limitada}\}$ . (c) Se  $f \geq 0$ ,  $f \in C_b(A)$ ,  $\mu_f$  denota a medida que tem  $f$  como derivada de Radon-Nikodym em relação a  $\lambda$ , i.e.,  $d\mu_f = f d\lambda$ . (d)  $L_1(A) = \{f: A \rightarrow \mathbb{R} \mid f \text{ é Lebesgue integrável, mensurável}\}$ ,  $\|f\|_1 = \int_A |f| d\lambda$  se  $f \in L_1(A)$ . (e)  $\mathcal{B}(A)$  denota a  $\sigma$ -álgebra de Borel de  $A$ .



Teorema 3.9. [Pianigiani, Yorke (1979)]

- (a) Seja  $T: \bar{A} \rightarrow \mathbb{R}^d$  expansora. Então existe  $\mu$  medida de probabilidade em  $\mathcal{B}(A)$  condicionalmente invariante e absolutamente contínua com relação a  $\lambda$ . No caso de  $T$  ser  $C^\infty$ ,  $\frac{d\mu}{d\lambda}$  tem versão  $C^\infty$ .
- (b) Se, ademais,  $T$  for transitiva nas componentes então existe uma única  $f \in K \stackrel{\text{def}}{=} \{g \in C_b(A) : g > 0, \int_A g d\lambda = 1\}$  tal que  $\mu_f$  seja condicionalmente invariante. (i.e., há uma única probabilidade condicionalmente invariante absolutamente contínua com densidade positiva, contínua e limitada.)
- (c) Se  $T$  for expansora e exata e se  $X$  for variável aleatória com lei  $\mu_g$ , onde  $g \in K$ , e  $\mu_m$  for a lei condicional de  $T^m X$  dado  $T^k X \in A$  para  $k = 1, \dots, m$ , então  $\mu_m$  converge no sentido  $w^*$  à probabilidade descrita em (b).

Exemplo. Uma classe bem simples de exemplos aos quais todos os resultados acima se aplicam é o das chamadas transformações de Rényi, definidas pelas propriedades: (Fig. 3.4)

$$T: [0, 1] \rightarrow [0, 1]$$

e existe partição  $0 = a_0 < a_1 < \dots < a_p = 1$  tal que se  $T_i = T$  restrita a  $(a_{i-1}, a_i)$ ,  $i = 1, \dots, p$ , então

(1)  $T_i$  tem extensão  $C^2$  a  $[a_{i-1}, a_i]$

(2)  $T_i [a_{i-1}, a_i] = [0, 1]$

$$(3) \quad \inf_{x \in (a_{i-1}, a_i)} |T_i'(x)| > 1$$

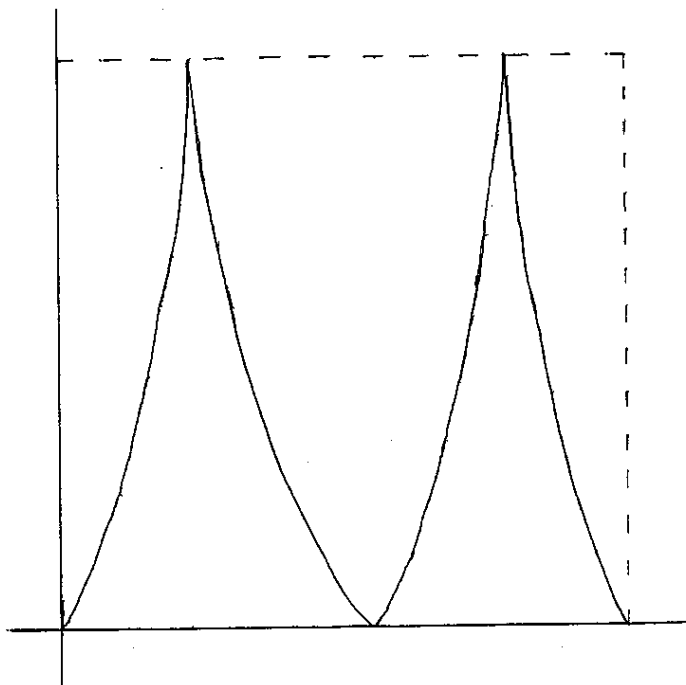


Fig. (3.4)

Uma tal  $T$  é expansora, exata, e para todo  $n \geq 1$ ,  $T^n$  também é expansora. Valem pois os itens (a), (b) e (c) do Teorema 3.9 (Ver tb. Lasota, Yorke (1981)). Neste caso  $\bar{A} = [0,1]$  e  $T[0,1] = [0,1]$  e a questão da saída de  $A$  não aparece; a medida em questão é realmente invariante e o resultado poderia ser deduzido de teoremas bem mais gerais.

Se  $T_\epsilon = (1+\epsilon)T$   $\epsilon > 0$  então as  $T_\epsilon$  ainda verificam todas as condições do Teorema 3.9. Neste caso teremos que  $[0,1] \setminus T_\epsilon^{-1}[0,1]$  é uma união de intervalos  $I_1^\epsilon, \dots, I_p^\epsilon$  cujos comprimentos são de ordem  $\epsilon$ . Neste caso pode-se ver que sendo

$$\tau_\epsilon(x) = \inf\{n \geq 1: T_\epsilon^n(x) \notin [0,1]\},$$

para  $x \in [0,1]$ , então existe  $\sigma > 0$  tal que para toda  $g \in K$

$$\mu_g\{x: \tau_\epsilon(x) > t\epsilon^{-1}\} \rightarrow e^{-\sigma t}$$

para todo  $t > 0$  i.e., se  $X$  tem lei  $\mu_0$ ,  $\tau_\epsilon(X)/\epsilon$  converge em lei à uma variável exponencial com média  $\sigma^{-1}$ . Isto será consequência do teorema a seguir enunciado.

Teorema 3.10. (Lasota, Yorke (1981))

Seja  $T_\epsilon: A \rightarrow \mathbb{R}^d$ ,  $\epsilon \in [0,1]$  uma família de transformações expansoras, exatas e tais que:

- (i) Pode-se tomar um mesmo  $\alpha > 1$  na Definição 3.6 para todas as  $T_\epsilon$ ,  $\epsilon \in [0,1]$ .
- (ii)  $(x, \epsilon) \rightarrow T_\epsilon(x)$ ;  $(x, \epsilon) \rightarrow DT_\epsilon(x)$  são uniformemente contínuas, e

$$\sup_{x, \epsilon} |D^2 T_\epsilon(x)| \leq \beta$$

para certo  $\beta < +\infty$ .

Seja  $\mu_0 = f_0 d\lambda$  a probabilidade condicionalmente invariante para  $T_0$ , descrita no ítem (b) do Teorema 3.9. Se

$$\frac{\mu_o(A \setminus T_\epsilon^{-1}(A))}{\epsilon} \longrightarrow \sigma \in (0, +\infty) \quad (3.14)$$

então para toda  $f \in K$

$$\mu_f[x: \tau_\epsilon(x) > t/\epsilon] \longrightarrow e^{-\sigma t} \quad (3.15)$$

para todo  $t > 0$ , onde

$$\begin{aligned} \tau_\epsilon(x) &= \inf\{n \geq 1: T_\epsilon^n x \in A\} \\ &= +\infty \text{ se } \{ \cdot \} = +\infty \end{aligned} \quad (3.16)$$

Comentários. Nesta breve exposição omitiremos as demonstrações dos teoremas acima enunciados, limitando-nos apenas a alguns comentários que julgamos convenientes: Inicialmente salientemos que a prova dos itens (a) e (b) no Teorema 3.9 (cf. Pianigiani, Yorke (1979)) não é construtiva, valendo-se do Teorema de ponto fixo de Schauder (em (a)). Por outro lado, vemos no Teorema 3.10 a importância de conhecer  $\mu_o(T_\epsilon^{-1} A)$ , para que se possa estimar o tempo de duração do "transiente" como mostram (3.14) e (3.15). Neste sentido, referimos a Pelikan (1985) para outra caracterização da "taxa de escape" do conjunto  $A$  ("quase atrator"). Os resultados de Pianigiani e Yorke (1979) e Lasota e Yorke (1981) resumidos nos Teoremas 3.9 e 3.10 acima, ainda são bem restritos. Pelo menos no caso de sistemas que satisfazem o Axioma  $A$ , os resultados de Bowen e Ruelle (1975) sugerem generalizações, onde conjuntos básicos desempenham o papel de "quase-atratores" (Ver

também Grebogi, Ott e Yorke (1983) e a secção §4.5 de Eckmann Ruelle (1985).) Em situações mais gerais (sem Axioma A) há apenas conjecturas e resultados numéricos (Kantz, Grassberger (1984)).

Uma questão importante além do "tempo de escape" refere-se ao comportamento das médias temporais quando o sistema ainda não deixou o "quase-atrator", e.g., no Teorema 3.10 são elas descritas por  $\mu_0$  ou medidas condicionalmente invariante? Isto representaria o outro ponto (referente à estabilidade das médias temporais) em um teorema análogo ao Teorema 3.3.

Referências adicionais:

Recentemente M. Cassandro, E. Olivieri e P. Picco (1985) estudam uma versão infinito dimensional do modelo considerado em §3.1 e §3.2. Nesta linha há muitos problemas interessantes; em E. Presutti (1985) há uma resenha do atual estado da pesquisa neste tópico bem como referências mais completas.

Os problemas discutidos em §3.1 e §3.2 para perturbações estocásticas de difeomorfismos, A. Galves, R. Langevin e M. E. Vares têm algumas respostas.

Referências

1. Azencott, R. (1980). Grandes Deviations et applications. École d'été de Saint Flour VIII-1978. Lecture Notes in Math., 774, Springer.
2. Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38 303-324.
3. Bahadur, R.R. (1971). Some Limit Theorems in Statistics SIAM. Philadelphia.
4. Bahadur, R.R. e Zabell, S.L. (1979). Large deviations of the sample mean in general vector spaces. Ann. Probability 7, 587-621.
5. Billingsley, P. (1965). Ergodic Theory and Information. Wiley- New York.
6. Binder, K. e Müller-Krumbhaar, H. (1974). Investigation of metastable states and nucleation in the Kinetic Ising model. Physical Review B6, 2328.
7. Bowen, R. e Ruelle, D. (1975). The ergodic theory of Axiom A flows. Inventiones Math. 29, 181-202.
8. Capocaccia, D., Cassandro, M., Olivieri, E. (1977). A study of metastability in the Ising model. Commun.Math.Phys., 39, 185-205

9. Cassandro, M., Galves, A., Olivieri, E., Vares, M.E. (1984) Metastable behavior of stochastic dynamics: a pathwise approach. *Journal of Statistical Physics*, 35, 603-634.
10. Cassandro, M., Olivieri, E., Picco, P. (em preparação).
11. Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations. *Ann. Math. Statist.* 23, 493-507.
12. Cramèr, H. (1937). Sur un nouveau théorème limite de la théorie des probabilités. *Colloquium on theory of probability*. Paris Hermann.
13. Day, M. (1982). Exponential levelling of stochastically perturbed dynamical systems, *SIAM J. Math. Anal.*, 13, 532-540.
14. Day, M. (1983). On the exponential exit law in the small parameter exit problem. *Stochastics*, 8, 297-323.
15. De Masi, A., Ianiro, N., Pellegrinotti, A., Presutti, E. (1984) A survey of the hydrodynamical behavior of many-particle systems. In: "Studies in Statistical Mechanics" 11, ed. J.L. Lebowitz and E.W. Montroll, North Holland.
16. Donsker, M. e Varadhan, S. (1975). Large deviations of Markov processes and the asymptotic evaluation of certain Markov process expectations for large time. *Lec. Notes in Math.* 451, 82-87. Springer.

17. Donsker, M., e Varadhan, S. (1975-76). Asymptotic evaluation of certain Markov process expectations for large times. I, II e III. *Comm. Pure App. Math.* 28, 1-47; 29, 279-301; 29, 389-461.
18. Eckmann, J.P., e Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. Preprint IHES.
19. Freidlin, M.I., e Wentzell, A.D. (1970). On small random perturbations of dynamical systems. *Russian Math. Surveys* 25, 1-55.
20. Freidlin, M.I., e Wentzell, A.D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.
21. Galves, A., Olivieri, E., Vares, M.E. (1984) Metastability for a class of dynamical systems subject to small random perturbations. Preprint IHES. (Submetido a *Ann. Probability*).
22. Grebogi, C., Ott, E., e Yorke, J.A. (1983). Crises, sudden changes in chaotic attractors and transient chaos. *Physica* 7D, 181-200.
23. Griffeath, D. (1979). Additive and cancelative interacting particle systems. *Lec. Notes in Math.* 724, Springer.
24. Groeneboom, P., Oosterhoff, J. e Ruyngaert, F.H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probability* 7, 553-586.



25. Kantz, H. e Grassberger P. (1984) Repellers, semi-attractors and long-lived chaotic transients. Preprint Wuppertal.
26. Kaplan, J.L. e Yorke, J.A. (1979) Perturbulence: a regime observed in a fluid flow model of Lorenz, Commun. Math. Phys. 67, 93-108.
27. Lanford, O. (1971). Entropy and equilibrium states in classical statistical mechanics. Lec. Notes in Phys. 20, 1-113.
28. Lasota, A. e Yorke, J.A. (1981) The law of exponential decay for expanding mappings. Rend. Den Mat. Univ. Padova, 64, 141-157.
29. Lebowitz, J.L. e Penrose, O. (1979). Towards a rigorous molecular theory of metastability. In "Fluctuation Phenomena", Eds. E.W. Montroll e S.L. Lebowitz. North Holland Publishing Company.
30. Li, T. e Yorke, J.A. (1975) Period three implies chaos. Amer. Math. Monthly 82, 985-992.
31. Lorenz, E.N. (1963). Deterministic non periodic flow. J. Atmos. Sci. 20, 130-141.
32. Mardsen, J.E. e McCracken, M. (1976) The Hopf Bifurcation and its Applications. Springer N.Y.
33. Pelikan, S. (1985). The duration of transients. Trans. Amer. Math. Soc. 287, 215-221.

34. Pianigiani, G. e Yorke, J.A. (1979). Expanding maps on sets which are almost invariant: decay and chaos. Trans.Amer. Math.Soc. 252, 351-366.
35. Presutti, E. (1985). Collective Phenomena in Stochastic Particle Systems. Preprint Univ. de Roma.
36. Robbins,K. A new approach to subcritical instability and turbulent transitions in a simple dynamo. Math. Proc. Cambridge Phil.Soc. 82, 309-325
37. Sanov, I. (1961). On the probability of large deviations of random variables. Selected trans.Math.Stat.Prob. I, 213-244.
38. Schonmann, Roberto H. (1984). Metaestabilidade para o processo de contacto: extensão dos teoremas básicos e estudo das Flutuações. Tese de doutoramento IME-USP. (aparecerá em Ann. Probability).
39. Serfling, R.J. (1980). Approximation Theorems in Mathematical Statistics. Wiley NY.
40. Williams, D. (1979). Diffusions, Markov Processes and Martingales. Wiley.

© 2000 by Pearson Education, Inc.



Impresso na Gráfica do

