INSTITUTO DE CIÊNCIAS MATEMÁTICAS DE SÃO CARLOS
INSTITUTO DE FÍSICA E QUÍMICA DE SÃO CARLOS
ESCOLA DE ENGENHARIA DE SÃO CARLOS

CONSIDERAÇÕES SOBRE OS MÉTODOS DE APROXIMAÇÃO ESTOCÁSTICA DE ROBBINS - MONRO E DE KIEFER - WOLFOWITZ

ÁREA DE PÓS-GRADUAÇÃO INTERUNIDADES
"CIÊNCIAS DE COMPUTAÇÃO E ESTATÍSTICA"

UNIVERSIDADE DE SÃO PAULO

SÃO CARLOS - SÃO PAULO
BRASIL

UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE CIÊNCIAS MATEMÁTICAS DE SÃO CARLOS

CONSIDERAÇÕES SOBRE OS MÉTODOS DE APROXIMAÇÃO ESTOCÁSTICA DE ROBBINS - MONRO E DE KIEFER - WOLFOWITZ

GERALDO GARCIA DUARTE JUNIOR

ORIENTADOR :

Prof. Dr. EUCLYDES C. DE LIMA FILHO

Dissertação apresentada ao Inst.Ciências Matemáticas de São Carlos da Universidade de São Paulo, na Área de Pós - Graduação Interunidades "Ciências de Computação e Estatística" para obtenção do título de Mestre .

São Carlos

1979

AGRADECIMENTOS

Agradeço de forma especial o Prof.Dr. Euclydes Custódio de Lima Filho pela ajuda e dedicação em nossa formação e a quem devemos nosso início na pós-graduação.

Aos Professores Antonio Acra Freiria e José Geraldo dos Reis pela grande ajuda no desenvolvimento dos cursos e estimulantes incentivos.

Aos colegas do Departamento de Geologia, Mineralogia, Física e Matemática da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, do Departamento de Genética e Matemática Aplicada a Biologia da Faculdade de Medicina de Ribeirão Preto e do Instituto de Ciências Matemáticas de São Carlos.

A Maria Antonieta V. B. Acra e Marisa Vellozo Losz Duarte pelo esmero e dedicação com a qual datilografaram esta dissertação.

Enfim, agradeço a todos que direta ou indiretamente colaboraram para a realização deste trabalho.

Este trabalho dependeu parcialmente do auxílio concedido pela CAPES , através de uma bolsa de pós-graduação

SUMMARY

GERALDO GARCIA DUARTE JUNIOR

Adviser:

Prof.Dr. Euclydes C. de Lima Filho

The purpose of this paper is to present the general ideas of Robbins-Monro's method of stochastic approximation .

The problem is the following :

Let $M(x)$ denot the expected value at level x of the response to a certain experiment . $M(x)$ is unknown to the experimenter, and it is desired to find the solution $x = \Theta$ of the equation $M(x) = \alpha$, where α is a given constant.

Robbins-Monro presents a method to do experiments at level $x_1, \ldots, x_n$ and such that x_n convergs to Θ with probability one .

I N D I C E

INTRODUÇÃO 1

CAPÍTULO I - APROXIMAÇÃO ESTOCÁSTICA 4

1 - O Problema na Forma Binária 4

2 - O Processo de Robbins-Monro 11

3 - Convergência com Probabilidade Um .. 22

CAPÍTULO II - MÁXIMO DE UMA FUNÇÃO DE REGRESSÃO .. 31

1 - O Método de Kiefer-Wolfowitz 31

2 - Convergência com Probabilidade Um .. 40

CAPÍTULO III - APLICAÇÕES 48

1 - Uma Aplicação em Biologia 48

2 - Um Modelo Teórico 52

3 - Máximo de uma Função de Regressão .. 59

4 - Comentários sobre os Processos 62

BIBLIOGRAFIA 64

INTRODUÇÃO

As idêias de aproximação estocástica começaram a ser discutidas na década de quarenta, veja por exemplo H. HOTELLING [6] e FRIEDMAN e SAVAGE [4], mas foi sômente em 1951 que H.ROBBINS e S. MONRO [9] deram um tratamento matemático ao problema. Esta formulação propiciou um desenvolvimento rápido e muitos artigos foram publicados, obtendo-se assim uma boa delimitação do problema.

Podemos formalizar o problema da sequinte maneira: suponhamos um experimento que a cada número real x, esteja associado uma variável aleatória $Y(x)$. Por exemplo: x é uma dose e $Y(x)$ é o efeito. Estamos procurando uma dose Θ que, em média, cause um efeito quantitativo α.

Nosso problema resume-se então, em achar a solução da equação $M(x) = \alpha$ quando $M(x) = E(Y(x))$. Uma forma de abordar o problema é a seguinte: Tomamos uma sequência $\{a_n\}$ de números reais positivos que seja decrescente e um valor inicial x_1. Conduzimos o experimento e observamos $y(x_1) = y_1$ da variável aleatória $Y(x_1)$ determinamos x_2 por $x_2 = x_1 + a_1(\alpha - y_1)$. Com valor x_2 achado, conduzimos o experimento e observamos $y(x_2) = y_2$ da variável aleatória $Y(x_2)$. Determinamos x_3 por $x_3 = x_2 + a_2(\alpha - y_2)$. Continuando deste modo, definimos uma sequência $\{x_n\}$ por :

$$x_{n+1} = x_n + a_n(\alpha - y_n)$$

a qual, segundo demonstração feita por ROBBINS - MONRO, converge em probabilidade para Θ, solução da equação $M(x) = \alpha$.

Tal técnica, mostra-se muito parecida com o método de NEWTON - RAPHSON para resolver a equação $M(x) = \alpha$, que consiste em escolher um valor x_1 arbitrário em um intervalo I e definir a sequência $\{ x_n \}$ por

$$x_{n+1} = x_n - (M'(x_n))^{-1} (M(x_n) - \alpha)$$

onde $M'(x_n)$ é a derivada de M no ponto $x = x_n$. Esta sequência converge para a solução desejada.

Como no método de NEWTON - RAPHSON, o processo de ROBBINS - MONRO tem também condições para que haja convergência. Estas condições serão discutidas em nosso trabalho, bem como uma importante consequência deste tratamento, que é o problema para determinar o máximo de uma função de regressão e foi estudado por KIEFER - WOLFOWITZ [7]. Podemos citar alguns casos em que tais processos podem ser aplicados. Por exemplo :

a) Sabemos que a dureza de uma liga é influenciada pelo tempoem que ela é exposta a uma certa temperatura. Seja x este tempo e $Y(x)$ a dureza da liga. O problema é determinar um tempo x para que a liga tenha uma dureza α .

b) Consideremos a sensibilidade de um explosivo a um choque. Cada explosivo tem um ponto crítico de choque. O problema será determinar este ponto.

c) Consideremos um lote de terra no qual x quilos de fertilizantes são aplicados e $Y(x)$ quilos de milho são produzidos. Provàvelmente quantidade menor de milho será colhida se a quantidade de fertilizantes for menor mas, um excesso de fertilizante poderá produzir uma quantidade menor de milho. O problema será determinar uma quantidade x para que se obtenha uma colheita máxima.

Nosso trabalho consistirá do estudo do processo de ROBBINS - MONRO, tendo em vista as aplicações. Para tanto, procuramos um desenvolvimento que tenha condições de fácil manuseio e não seja muito restrito. No estudo de máximo de uma função de regressão temos também hipóteses que facilitam as aplicações. No último capítulo, criamos modelos e aplicamos o processo fazendo algumas observações.

CAPÍTULO I

1. *O Problema na Forma Binária*

Para uma primeira abordagem do problema de aproximação estocástica, estudaremos aqui o problema do tipo resposta-não-resposta. Alguns experimentos são desta forma, por exemplo quando testamos um inseticida. Neste caso aplicamos uma dose de inseticida e observamos se o inseto morre ou não. No caso de morte, obtemos uma resposta. O problema então será determinar uma dose crítica para uma dada resposta quantitativa.

Matematicamente o problema pode ser assim formulado:

Seja Z uma variável aleatória com função distribuição $M(z)$. Para cada número real x, seja $Y(x)$ uma variável aleatória tal que

$$Y(x) = 1 \quad \text{se} \quad Z \leqslant x \qquad \text{(Resposta)}$$

$$Y(x) = 0 \quad \text{se} \quad Z > x \qquad \text{(não Resposta)}$$

Podemos associar x a dose e $Y(x)$ a resposta observada.

Temos então :

$$P\,[Y(x) = 1] = P\,[Z \leqslant x] = M(x)$$

$$P\,[Y(x) = 0] = P\,[Z > x] = 1 - P\,[Z \leqslant x] = 1-M(x)$$

Logo

$$E\,[Y(x)] = 1.M(x) + 0(1-M(x)) = M(x)$$

O problema está bem formalizado no seguinte teorema:

<u>*TEOREMA 1.1*</u> : Seja M uma função distribuição e α um número real tal que exista um número real Θ com $M(\Theta) = \alpha$. Seja $M(x)$ diferenciável em Θ com $M'(\Theta) > 0$. Seja x um número real, n um inteiro positivo e

$$X_{n+1} = x_n + \frac{1}{n}(\alpha - Y_n) \qquad (Y_n = Y(x_n))$$

quando Y_n é tal que

$$P\left[Y_n = 1/X_1, \ldots, X_n, Y_n, \ldots, Y_{n-1} \right] = M(X_n)$$

$$P\left[Y_n = 0/X_1, \ldots, X_n, Y_1, \ldots, Y_{n-1} \right] = 1 - M(X_n)$$

Então

$\lim\limits_{n\to\infty} E(X_n - \Theta)^2 = 0$ e portanto X_n converge para Θ em probabilidade.

<u>*Dem*</u> :- Seja $b_n = E(X_n - \Theta)^2$

Mostremos que $\lim\limits_{n\to\infty} b_n = 0$

Temos

$$X_{n+1} - \Theta = X_n - \Theta - \frac{1}{n}(Y_n - \alpha)$$

$$E(X_{n+1} - \Theta)^2 = E(X_n - \Theta)^2 - \frac{2}{n} E\{(X_n - \Theta)(Y_n - \alpha)\} + \frac{1}{n^2} E\{(Y_n - \alpha)^2\}$$

Consideremos

$$d_n = E\{(X_n - \Theta)(Y_n - \alpha)\}$$

$$e_n = E\{(Y_n - \alpha)^2\}$$

Então

$$b_{n+1} = b_n - \frac{2}{n} d_n + \frac{1}{n^2} e_n$$

$$(1) \quad \sum_{j=1}^{n} (b_{j+1} - b_j) = -2 \sum_{j=1}^{n} \frac{d_j}{j} + \sum_{j=1}^{n} \frac{e_j}{j^2}$$

Sendo

(2) $Y(x)$ Zero ou Um e $0 < \alpha < 1$

Temos

$$0 \leqslant e_n = E\{(Y_n - \alpha)^2\} \leqslant 1$$

$\sum_{j=1}^{n} \frac{e_j}{j^2}$ é crescente e limitada por $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$

Logo $\sum_{j=1}^{\infty} \frac{e_j}{j^2}$ é convergente.

(3) $d_n = E\{(X_n - \Theta)(Y_n - \alpha)\} =$

$= E\{E[(X_n - \Theta)(Y_n - \alpha) / X_1, \ldots, X_n, Y_1, \ldots, Y_{n-1}]\} =$

$= E\{(X_n - \Theta)(M(X_n) - \alpha)\} > 0$ pois $M(x)$ é crescente e $M(\Theta) = \alpha$

De (1) e (3) vemos que

$$2 \sum_{j=1}^{n} \frac{d_j}{j} \leqslant \sum_{j=1}^{n} \frac{e_j}{j^2} \leqslant \sum_{j=1}^{\infty} \frac{e_j}{j^2} < \infty$$

Logo $\sum_{j=1}^{\infty} \frac{d_j}{j}$ é convergente.

Portanto, como $b_{n+1} = b_1 - 2 \sum_{j=1}^{n} \frac{d_j}{j} + \sum_{j=1}^{n} \frac{e_j}{j^2}$

$\lim b_n = b_1 - 2 \sum_{1}^{\infty} \frac{d_n}{n} + \sum_{n=1}^{\infty} \frac{e_n}{n^2} = b$ existe.

Mostremos que $b = 0$

Seja

$$A_1 = |x_1 - \theta|$$

$$A_n = |x_1 - \theta| + \sum_{j=1}^{n-1} \frac{1}{j} \quad \text{se} \quad n > 1$$

Temos

$$X_{n+1} - X_n = - \frac{1}{n} (Y_n - \alpha)$$

Logo

$$X_{n+1} - X_1 = - \sum_{j=1}^{n} \frac{Y_j - \alpha}{j} \qquad \text{e}$$

$$|X_{n+1} - \theta| = |X_1 - \theta - \sum_{j=1}^{n} \frac{Y_j - \sigma}{j}| \leqslant |X_1 - \theta| + \sum_{j=1}^{n} \frac{1}{j} |Y_j - \alpha| \leqslant$$

$\leqslant |X_1 - \theta| + \sum_{j=1}^{n} \frac{1}{j}$ pois $|Y_j - \alpha| \leqslant 1$ para todo j conforme (2)

Então

$$|X_n - \theta| \leqslant A_n \qquad n > 1$$

Seja $k_n = \inf_{|x-\theta|<A_n} \{\frac{M(x)-\alpha}{x - \theta}\}$, $k_n > 0$

$$d_n = E\{(X_n - \theta)(Y_n - \alpha)\} = E\{(X_n - \theta)^2 \cdot \frac{M(X_n) - \alpha}{X_n - \theta}\} \geqslant$$

$$\geq E\{(X_n-\theta)^2\} \inf_{|x-\theta|\leq A_n} \left\{\frac{M(x)-\alpha}{x-\theta}\right\} = b_n k_n$$

Portanto

(4) $\quad d_n \geq b_n k_n$

(5) Sendo $\lim_{x\to\theta} \dfrac{M(x)-\alpha}{x-\theta} = \lim_{x\to\theta} \dfrac{M(x)-M(\theta)}{x-\theta} = M'(\theta) > 0$

existe $\delta > 0$ tal que $|x-\theta| < \delta \Longrightarrow \dfrac{M(x)-\alpha}{x-\theta} > \dfrac{1}{2} M'(\theta)$

Logo $\inf_{|x-\theta|<\delta} \left\{\dfrac{M(x)-\alpha}{x-\theta}\right\} > \dfrac{1}{2} M'(\theta)$

Como $\lim A_n = \infty$, existe N tal que se $n > N$ então $A_n > \delta$

Estudemos $\inf_{\delta\leq|x-\theta|\leq A_n} \left\{\dfrac{M(x)-\alpha}{x-\theta}\right\}$

Para $\theta + \delta \leq x \leq A_n + \theta$

$$\frac{M(x)-\alpha}{x-\theta} \geq \frac{M(\theta+\delta)-\alpha}{A_n}$$

De (5) temos $M(x) - \alpha \geq \dfrac{1}{2} M'(\theta)(x-\theta)$ p/ $|x-\theta| \leq \delta$

Quando $x = \theta + \delta$ temos

$$M(\theta+\delta) - \alpha > \frac{\delta}{2} M'(\theta)$$

Neste caso

$$\frac{M(x) - \alpha}{x - \theta} \geqslant \frac{\delta \; M'(\theta)}{2A_n}$$

Para $\theta - A_n \leqslant x \leqslant \theta - \delta$

$$\frac{M(x) - \alpha}{x - \theta} = \frac{\alpha - M(x)}{\theta - x} \geqslant \frac{\alpha - M(\theta-\delta)}{A_n}$$

De (5) temos $\dfrac{\alpha - M(x)}{\theta - x} \geqslant \dfrac{1}{2} M'(\theta)$

Quando $x = \theta - \delta$ temos $\alpha - M(\theta - \delta) \geqslant \dfrac{\delta}{2} M'(\theta)$

Neste caso também vale a desigualdade

$$\frac{M(x) - \alpha}{x - \theta} \geqslant \frac{\delta \; M'(\theta)}{2A_n}$$

Como $\dfrac{\delta}{A_n} \leqslant 1$

$$k_n = \inf_{|x-\theta| < A_n} \left\{ \frac{M(x) - \alpha}{x - \theta} \right\} \geqslant \frac{\delta \; M'(\theta)}{2A_n} = \frac{k}{A_n} \quad \text{para } n > N$$

Então

$$\sum_{1}^{\infty} \frac{k_n}{n} \geqslant \sum^{\infty} \frac{1}{nA_n} \geqslant \frac{k}{k_1} \sum^{\infty} \frac{1}{n \;\ln\; n} = \infty$$

Pois $\ln\, n = \sum_{1}^{n-1} \int_{j}^{j+1} \dfrac{dt}{t} \geqslant \sum_{1}^{n-1} \dfrac{1}{j+1} = \sum_{j=2}^{n} \dfrac{1}{j}$ e

$$A_n = |x_1 - \theta| + \sum_{j=1}^{n-1} \frac{1}{j} \leqslant |x_1 \pm \theta| + 1 + \sum_{j=2}^{n} \frac{1}{j} \leqslant$$

$$\leqslant (|x_1 - \theta| + 1) \sum_{j=2}^{n} \frac{1}{j} + \sum_{j=2}^{n} \frac{1}{j} \leqslant$$

$$< (|x_1 - \theta| + 2) \sum_{j=1}^{n} \frac{1}{j} < (|x_1 - \theta| + 2) \ln n =$$

$$= k_1 \ln n$$

Determinamos assim, uma sequência k_n tal que

a) $k_n \geqslant 0$

b) $d_n \geqslant k_n b_n$

c) $\sum \frac{k_n}{n} = \infty$

Então

$$(6) \qquad 0 \leqslant \sum \frac{1}{n} k_n b_n \leqslant \sum \frac{d_n}{n} < \infty$$

Supondo $b > 0$ existe N tal que se $n > N$ então $b_n > \frac{b}{2}$

Logo

$$\sum_{N}^{\infty} \frac{1}{n} k_n b_n > \frac{b}{2} \sum_{N}^{\infty} \frac{k_n}{n} = \infty$$

contrariando (6)

Portanto $b = 0$

Este é o caso mais simples pois, a variável aleatória $Y(x)$ tem massa somente nos pontos Zero e Um.

Não precisamos exigir que $M(x)$ seja diferenciável . Veremos mais adiante, em casos mais gerais, que podemos substituir esta hipótese por outras mais fracas.

2. *O Processo de ROBBINS - MONRO*

Suponhamos que a cada valor real x esteja associada uma variável aleatória $Y = Y(x)$ com função distribuição

$$P\left[\, Y(x) \leqslant y \,\right] = H\,(y/x) \qquad \text{tal que}$$

$$M(x) = \int_{\infty}^{\infty} y\,dH(\,y/x\,) \quad , \quad P\left[\,|Y(x)| \leqslant C\,\right] =$$

$$= \int_{-C}^{C} dH(\,y/x\,) = 1 \quad \text{e que} \quad M(x) = \alpha \quad \text{tenha uma única raiz } \Theta$$

Definimos a sequência $\{\, X_n \,\}$ por

$$X_{n+1} = X_n - a_n\,(\,Y_n - \alpha\,)$$

onde Y_n é uma variável aleatória tal que

$$P\,[Y_n \leqslant y/x_n\,] = H\,(y/x_1,\ldots,x_n,\,y_1,\ldots,y_{n-1}) = H\,(y/x_n)$$

Mostremos que, sob certas condições, a sequência $\{\, X_n \,\}$ converge para Θ em probabilidade

Seja

$$b_n = E\,(\,X_n - \Theta\,)^2 > 0$$

Sabemos que se $\lim b_n = 0$ então x_n converge para θ em probabilidade.

$$b_{n+1} = E\,(X_{n+1} - \theta)^2 = E\,\{E\,[(X_{n+1} - \theta)^2 / x_n]\} =$$

$$= E\,\{E\,[(X_n - \theta - a_n(Y_n - \alpha))^2 / x_n]\} =$$

$$= b_n - 2a_n\, E\,\{(X_n - \theta)(M(X_n) - \alpha)\} + a_n^2\, E\,\{\int_{\infty}^{\infty} (y-\alpha)^2\, dH(y/x_n)\}$$

Tomando

$$d_n = E\,\{(X_n - \theta)\,(M(X_n) - \alpha)\}$$

$$e_n = E\,\{\int_{\infty}^{\infty} (y - \alpha)^2\, dH\,(y/x_n)\}$$

Temos

$$b_{n+1} - b_n = a_n^2\, e_n - 2a_n\, d_n$$

Logo

$$b_{n+1} = b_n - 2\sum_{i=1}^{n} a_i\, d_i + \sum_{i=1}^{n} a_i^2\, e_i \tag{1.1}$$

Suponhamos que

(A) $M(x) < \alpha$ para $x < \theta$ e $M(x) > \alpha$ para $x > \theta$

(B) $\{a_n\}$ uma sequência positiva tal que

$$\sum_{n=1}^{\infty} a_n = \infty \quad \text{e} \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$

(C) $\quad k_n = \inf \{ \dfrac{M(x) - \alpha}{x - \theta} \}$ para $0 < | x - \theta| \leqslant A_n$

onde $\quad A_n = |x_1 - \theta | + (C + |\alpha |)(a_1 + \ldots + a_{n-1}) \quad n > 1$

$$A_1 = |x_1 - \theta |$$

(D) $\quad k_n > \dfrac{k}{A_n}$ para algum $k > 0$ e n suficientemente grande

Com estas condições

$$0 \leqslant e_n = E \{ \int_{\infty}^{\infty} (y - \alpha)^2 \, dH(y/x_n) \} \leqslant$$

$$\leqslant E \int_{\infty}^{\infty} (|y| + |\alpha|)^2 \, dH(y/x_n) = E \{ \int_{-C}^{C} (|y| + |\alpha|)^2 \, dH(y/x_n) \} \leqslant$$

$$\leqslant E \{ \int_{-C}^{C} (C + |\alpha|)^2 \, dH(y/x_n) \} \leqslant (C + | \alpha |)^2$$

Logo

$$(1.2) \qquad 0 \leqslant \sum_{n=1}^{\infty} a_n^2 \, e_n \leqslant \sum_{n=1}^{\infty} a_n^2 (C + | \alpha |)^2 < \infty$$

De (1.1) e (A)

$$d_n \geqslant 0$$

$$0 \leqslant 2 \sum_{i=1}^{n} a_i \, d_i \leqslant b_1 + \sum_{i=1}^{n} a_i^2 \, e_i \leqslant b_1 + \sum_{i=1}^{\infty} a_i^2 \, e_i < \infty$$

Assim

(1.3) $\qquad \Sigma \, a_n \, d_n$ é convergente

Então

$$\lim_{n\to\infty} b_n = b_1 - 2 \sum_{n=1}^{\infty} a_n d_n + \sum_{n=1}^{\infty} a_n^2 e_n = b \text{ existe}$$

Mostremos que $b = 0$

$$|X_n - \theta| = |(x_1 - \theta) - \sum_{i=1}^{n-1} a_i (Y_i - \alpha)| \leqslant$$

$$\leqslant |x_1 - \theta| + \sum_{i=1}^{n-1} a_i (|Y_i| + |\alpha|)$$

Daí vemos que

$$P\{|X_n - \theta| \leqslant A_n\} \geqslant P\{|x_1 - \theta| + \sum_{i=1}^{n-1} a_i (|Y_i| + |\alpha|) \leqslant A_n\} =$$

$$= P\{\sum_{i=1}^{n-1} a_i |Y_i| \leqslant c \sum_{i=1}^{n-1} a_i\}$$

Sendo que $P\{|Y_i| \leqslant C\} = 1$

Temos

$$P\{\sum_{i=1}^{n-1} a_i |Y_i| \leqslant C \sum_{i=1}^{n-1} a_i\} = 1$$

Portanto

$$P\{|X_n - \theta| \leqslant A_n\} = 1$$

Vemos então que

$$d_n = E\{(X_n - \theta)(M(X_n) - \alpha)\} = E\{(X_n - \theta)^2 \frac{M(X_n) - \alpha}{X_n - \theta}\} \geqslant$$

$$\geqslant k_n E\{(X_n - \theta)^2\} = k_n b_n$$

Para n suficientemente grande

$$2\ (C + |\alpha|)\ (a_i + \dots + a_{n-1}) \geq A_n$$

Então por (C)

$$(1.4) \quad \Sigma\, a_n\, k_n \geq \Sigma\, \frac{a_n\, k}{A_n} \geq \frac{k}{2(C+|\alpha|)}\ \Sigma\, \frac{a_n}{a_1+\dots+a_{n-1}} = \infty$$

A sequência k_n satisfaz

$$(K_1) \quad \Sigma\, a_n\, k_n = \infty$$

$$(K_2) \quad d_n \geq k_n\, b_n$$

De (K_2) e (1.3)

$$(K_3) \quad \Sigma\, a_n\, k_n\, b_n < \infty$$

(1.5) Se $b > 0$ então existe N tal que

$$n > N \qquad b_n > \frac{1}{2}\, b$$

Por (1.4) e (1.5)

$$\sum_{n=N}^{\infty} a_n\, k_n\, b_n > \sum_{n=N}^{\infty} \frac{1}{2}\, b\ a_n\, k_n = \infty$$ que contradiz (K_3)

Portanto $b = 0$

As hipóteses (D) e (A) podem ser substituidas por

$$M(x) < \alpha - \delta \qquad \text{para} \quad x < \Theta$$

$$M(x) > \alpha - \delta \qquad \text{para} \quad x > \Theta$$

Esta hipótese é mais forte que (A) e também garante que $k_n > \frac{\delta}{A_n}$ que é exigido em (D) . Com estas hipóteses, nada é mudado na demonstração e portanto X_n converge para Θ em probabilidade.

Se $M(x)$ satisfaz :

$M(x)$ é crescente

$M(\Theta) = \alpha$

$M'(\Theta) > 0$

Então temos

$$M(x) < \alpha \quad p/\ x < \Theta \quad e \quad M(x) > \alpha \quad p/\ x > \Theta$$

Para algum $\delta > 0$ e n suficientemente grande

$$k_n > \frac{\delta\, M'(\Theta)}{2A_n} = \frac{k}{A_n}$$

Neste caso também podemos substituir (D) e (A) por esta nova hipótese que continuaremos tendo a convergência de X_n para Θ em probabilidade.

A hipótese $P\{\,|Y(x)| \leq C\,\} = 1$, tomada na demonstração da convergência, pode ser substituida por

$$|M(x)| \leq C < \infty\,, \quad \int_{\infty}^{\infty} (y - M(x))^2\, dH(y/x) \leq \sigma^2$$

como sugere ROBBINS - MONRO [9] .

WOLFOWITZ resolveu este problema da seguinte maneira :

Consideremos :

(2.1) $|M(x)| \leqslant C < \infty$, $\int_{\infty}^{\infty} (y-M(x))^2 \, dH(y/x) \leqslant \sigma^2 < \infty$

(2.2) $M(x) < \alpha - \delta$ para $x < \theta$

$M(x) > \alpha + \delta$ para $x > \theta$

(2.3) $M(x) < \alpha$ para $x < \theta$

$M(\theta) = \alpha$

$M(x) > \alpha$ para $x > \theta$

Para algum $\delta > 0$, $M(x)$ é estritamente crescente em $|x - \theta| < \delta$ e $\inf\limits_{|x-\theta|>\delta} \{ M(x) - \alpha \} > 0$

Definimos

$$X_{n+1} = X_n - a_n (Y_n - \alpha)$$

TEOREMA : Se (2.1) e (2.2) ou (2.1) e (2.3) valem, então X_n converge para θ em probabilidade.

Dem.:

Consideremos

$$b_n = E \{ (X_n - \theta)^2 \}$$

$$d_n = E \{ (X_n - \theta) (M(X_n) - \alpha) \}$$

$$e_n = E \{ \int_{\infty}^{\infty} (y - \alpha)^2 \, dH (y/x_n) \}$$

Então

$$0 \leqslant e_n = E \quad \{ \int_{-\infty}^{\infty} (y - \alpha)^2 \quad dH(y/x_n) \ \} \quad =$$

$$= E \ \{ \int_{-\infty}^{\infty} (y - M(x_n))^2 \quad dH(y/x_n) \ \} \quad +$$

$$+ \ 2 \ E \ \{ \int_{\infty}^{\infty} (y - M(x_n)) \ (M(x_n) - \alpha) \quad dH(y/x_n) \ \} \quad +$$

$$+ \ E \ \{ \int_{-\infty}^{\infty} (M(x_n) - \alpha)^2 \quad dH(y/x_n) \ \}$$

Sendo

$$M(x) \quad = \quad \int_{-\infty}^{\infty} \ y dH(y/x)$$

$$2 \ E \quad \{ \int_{-\infty}^{\infty} (y - M(x_n)) \quad (M(x_n) - \alpha) \quad .dH(y/x_n) \} \ = \ 0$$

Portanto

$$e_n \quad = \quad E \quad \{ \int_{-\infty}^{\infty} (y - M(x_n))^2 \quad dH(y/x_n) \ \} \ +$$

$$+ \ E \quad \{ \ (M(x_n) - \alpha)^2 \} \ \leqslant \ \sigma^2 \ + (C + | \ \alpha \ | \ )^2 \ = \ h$$

Com estas condições vimos que

$$\lim_{n\to\infty} \ b_n \ = \ b_1 + \ \Sigma \ a_n^2 \ e_n \ - \ 2 \ \Sigma \ a_n \ d_n \ = \ b \quad \text{existe}$$

Sendo

$$d_n \ \geqslant \ 0 \quad , \quad \Sigma \ a_n \ d_n \ < \ \infty \quad e \quad \Sigma \ a_n \ = \ \infty$$

Entaõ

$$\lim \ \inf \ d_n \ = \ 0$$

Seja $\{ d_{nj} \}$ uma subsequência de $\{ d_n \}$ tal que

(2.4) $\lim d_{nj} = 0$

Mostremos que $\{ X_{nj} \}$ converge para θ

Se isto não ocorrer, existirá uma subsequência $\{ t_j \}$ da sequên cia $\{ n_j \}$ e números positivos ε e η tal que para todo t_j temos

$$P \{ | X_{tj} - \theta | > \eta \} > \varepsilon$$

Mas para todo t_j temos

$$d_{tj} = E \{ (X_{tj} - \theta) (M (X_{tj}) - \alpha) \} =$$

$$= E \{ |X_{tj} - \theta | \, |M(X_{tj}) - \alpha | \} = \int_{-\infty}^{\infty} |x - \theta| \, | M(x) - \alpha | dP(X_{tj} \leqslant x) =$$

$$= \int_{|x-\theta|>\eta} | x - \theta| \, |M(x) - \alpha| \, dP(X_{tj} \leqslant x) +$$

$$+ \int_{|x-\theta|\leqslant\eta} | x - \theta| \, |M(x) - \alpha| \, dP(X_{tj} \leqslant x) \geqslant$$

$$\geqslant \eta \inf_{|x-\theta|>\eta} \{| M(x) - \alpha| \} \int_{|x -\theta|\geqslant\eta} dP(X_{tj} \leqslant x) \geqslant$$

$$\geqslant \varepsilon \, \eta \inf_{|x-\theta|>\eta} \{ | M(x) - \alpha | \}$$

Por (2.2) ou (2.3) $d_{tj} \geqslant \varepsilon \, \eta \inf_{|x-\theta|\geqslant\eta} \{|M(x) - \alpha| \} > 0$. Isto- contradiz (2.4) .

Logo X_{nj} converge para θ em probabilidade.

Dado ε e η números positivos arbitrários, tomamos s um número positivo tal que

$$(2.5) \qquad \frac{s^2 + s}{n^2} < \frac{\varepsilon}{2}$$

Como X_{nj} converge para θ , existe um inteiro positivo N_0 tal que

$$P \{ | X_{N_o} - \theta | > s \} < \frac{\varepsilon}{2}$$

Sendo que $\Sigma a_n^2 < \infty$, podemos escolher o N_o de modo que

$$\sum_{N_o}^{\infty} a_n^2 < \frac{s}{2h}$$

Definimos para $n > N_o$

$$b_n(x) = E \{ (X_n - \theta)^2 / X_{N_o} = x \}$$

Temos então

$$b_{n+1}(x) = E \{ E [(X_{n+1} - \theta)^2 / X_{N_o} = x, x_n] / X_{N_o} = x \} =$$

$$= E \{ E [(X_n - \theta)^2 / X_{N_o} = x, x_n] / X_{N_o} = x \} -$$

$$- 2 a_n E \{ E [(X_n - \theta) (Y_n - \alpha) / X_{N_o} = x, x_n] / X_{N_o} = x \} +$$

$$+ 2 a_n^2 E \{ E [(Y_n - \theta)^2 / X_{N_o} = x, x_n] / X_{N_o} = x \}$$

Logo

$$b_{n+1}(x) = E \{ (X_{N_o} - \theta)^2 / X_{N_o} = x \} -$$

$$- 2 \sum_{i=n_o}^{n} a_i E \{ (X_i - \theta) (M(X_i) - \alpha) / X_{N_o} = x \} +$$

$$+ \sum_{N_o}^{n} a_i^2 \; E \; \{ E \left[(Y_i - \theta)^2 / X_{N_o} = x, x_i \right] / X_{N_o} = x \} <$$

$$< (x - \theta)^2 + h \sum_{i=N_o}^{n} a_i^2 < (x - \theta)^2 + \frac{s}{2}$$

Seja $A = \{ \, | x - \theta | < s \, \}$, $A_{N_o} = X_{N_o}^{-1} (A)$

$$E_{A_{N_o}} \{ |X_n - \theta| \} = \frac{1}{P(A_{N_o})} \int_{A_{N_o}} | X_n - \theta | \, dP =$$

$$= \frac{1}{P(A_{N_o})} \int_{A} E \, (|X - \theta| / X_{N_o} = x) \, dP \, (X_{N_o} \leqslant x) \leqslant$$

$$\leqslant \frac{1}{P(A_{N_o})} \int_{A} ((x - \theta)^2 + \frac{s}{2}) \, dP \, (X_{N_o} \leqslant x) \leqslant$$

$$< \frac{1}{P(A_{N_o})} \int_{A} (s^2 + \frac{s}{2}) \, dP \, (X_{N_o} \leqslant x) =$$

$$= s^2 + \frac{s}{2} < s^2 + s$$

Pela desigualdade de TCHEBYCHEFF

$$(2.6) \quad P \{ |X_n - \theta| > \eta / |X_{N_o} - \theta| < s \} \leqslant \frac{E \, \{(X_n - \;)^2 / \; X_{N_o} - \; s \}}{\eta^2} < \frac{s^2 + s}{\eta^2}$$

Dado ε e η positivos, para $n > N_o$ temos

$$P \{ |X_n - \theta| > \eta \} =$$

$$= P \{ |X_{N_o} - \theta| < s \} \; P \{ |X_n - \theta| > \eta / |X_{N_o} - \theta| < s \} +$$

$$+ \quad P\{\,|X_{N_o} - \theta| > s\,\} \,.\, P\,\{|X_n - \theta\,| > \eta \,/\, |X_{N_o} - \theta| > s\,\}$$

Por (2.6) e (2.5) temos

$$P\,\{|\,X_n - \theta\,| > \eta\,\} < P\,\{\,|X_{N_o} - \theta| < s\,\}\,\frac{\varepsilon}{2} +$$

$$+ \frac{\varepsilon}{2}\,P\,\{\,|\,X_n - \theta| > \eta \,/\, |X_{N_o} - \theta| > s\,\} < \varepsilon$$

Mostrando que $\{ X_n \}$ converge para θ em probabilidade .

3. *Convergência com Probabilidade Um*

As convergências estudadas até aqui foram convergências em probabilidade. Este tipo de convergência é mais fraco que convergência com probabilidade um, isto é, convergência com probabilidade um implica convergência em probabilidade .

A demonstração de que, sob certas condições, a sequência $\{ X_n \}$ converge para θ com probabilidade um, foi feita por BLUM [2] com hipóteses bastante simples e que passaremos a estudar.

Lema 1 : Seja $\{ V_n \}$ uma sequência de variáveis aleatórias tal que $\Sigma\, E\,\{ V_n^{\,2} \} < \infty$. Então

$\sum_{j=1}^{n} [\, V_j - E\,(\, V_j \,/\, V_1 \,,\, \ldots \,,\, V_{j-1}\,)\,]$ converge para uma variável aleatória com probabilidade um.

Este lema foi demonstrado por LOÈVE [8].

Seja $M(x)$ uma função de regressão correspondente a familia $H(y/x)$. Suponhamos as seguintes condições :

(3.1) $| M(x) | < c + d |x|$

(3.2) $\int_{-\infty}^{\infty} (y - M(x))^2 \, dH(y/x) \leq \sigma^2$

(3.3) $M(x) < \alpha$ para $x < \Theta$, $M(x) > \alpha$ para $x > \Theta$

(3.4) $\inf_{\delta_1 \leq |x-\Theta| \leq \delta_2} \{| M(x) - \alpha|\} > 0$ para todo $0 < \delta_1 < \delta_2 < \infty$

Como anteriormente $\{ a_n \}$ é uma sequência real positiva tal que

$$\Sigma a_n = \infty , \qquad \Sigma a_n^2 < \infty$$

Tomamos

$$X_{n+1} = X_n - a_n (Y_n - \alpha)$$

Lema 2 : Se (3.2) vale, então a sequência

$$\{ X_{n+1} - \sum_{j=1}^{n} a_j (\alpha - M(X_j)) \}$$ converge para uma variável aleatória com probabilidade um.

Dem :

Seja $v_j = a_j (Y_j - M(X_j))$

$$E \{ v_j^2\} = a_j^2 E \{ (Y_j - M(X_j))^2\} \leq a_j^2 \sigma^2$$

Sendo que $\Sigma a_n^2 < \infty$ temos $\Sigma E \{ v_j^2 \} < \infty$

Mostremos que

$$E \{ (Y_n - M(X_n) / Y_1 - M(X_1) , \ldots , Y_{n-1} - M(X_{n-1}) \} = 0$$

Para isto observamos que dado x_1 uma constante temos $M(x_1)$. Mas dado $Y_1 - M(x_1)$ e $M(x_1)$ nós obtemos x_2 .. Continuando, obtemos

$$E \{ (Y_n - M(X_n)) / Y_1 - M(x_1), \ldots , Y_{n-1} - M(x_{n-1}) \} =$$

$$= E \{ (Y_n - M(X_n) / X_n \} = 0$$

Pelo lema 1

$\sum_{j=1}^{n} a_j (Y_j - M(x_j))$ converge para uma variável aleatória com probabilidade um.

Portanto

$$X_{n+1} - \sum_{j=1}^{n} a_j (\alpha - M(x_j)) = X_1 + \sum_{j=1}^{n} a_j (M(x_j) - Y_j)$$

converge para uma variável aleatória com probabilidade um.

Lema 3 : Se (3.1) , (3.2) e (3.3) valem, então $\{ X_n \}$ converge com probabilidade um.

Dem :

Mostremos primeiramente que

$$P \{ \lim x_n = \infty \} + P \{ \lim x_n = - \infty \} = 0$$

Suponhamos que $\{ x_n \}$ seja uma sequência amostral tal que $\lim x_n = \infty$

(Para $\{ x_n \}$ tal que $\lim x_n = -\infty$ a demonstração é semelhante)

Então existirá um N tal que $n > N$ temos $x_n > \theta$ e portanto

$$a_n (\alpha - M(x_n)) < 0 \quad \text{por (3.3)}$$

Mas então

$$\lim_{n \to \infty} (x_{n+1} - \sum_{j=1}^{n} a_j (\alpha - M(x_j)) = \infty$$

Conforme o lema 2 isto ocorre somente com probabilidade zero. Mostramos assim que

$$P \{ \lim x_1 = \infty \} + P \{ \lim x_n = -\infty \} = 0$$

Suponhamos agora, por absurdo, que o lema seja falso. Então existe um conjunto de sequências amostrais, com probabilidade positiva, satisfazendo :

(a) $x_{n+1} - \sum_{i=1}^{n} a_i (\alpha - M(x_i))$ converge para um número finito

(b) $\lim \inf x_n < \lim \sup x_n$

para toda sequência neste conjunto .

Seja $\{ x_n \}$ uma tal sequência e suponhamos que $\lim \sup x_n > \theta$ (os mesmos argumentos valem se $\lim \sup x_n > \theta$)

Escolhemos números a e b satisfazendo

$$a > \theta \ , \ \lim \inf x_n < a < b \leq \lim \sup x_n$$

Como $\lim a_n = 0$ e $x_{n+1} - \sum_{i=1}^{n} a_i\ (\alpha - M(x_i))$ converge para um número finito, escolhemos N tal que se $N < n < m$ então

$$\text{(3.5)} \quad \begin{aligned} &\text{a)} \quad a_n < \min \left\{\frac{1}{3d}\ ,\ \frac{b - a}{3(|\alpha|+c+d\ |\theta|)}\right\} \\ &\text{b)} \quad \left|x_m - x_n - \sum_{n}^{m-1} a_i\ (\alpha - M(x_i))\right| < \frac{b-a}{3} \end{aligned}$$

Escolhemos m e n tal que

$$\text{(3.6)} \quad \begin{aligned} &\text{a)} \quad N < n < m \\ &\text{b)} \quad x_n < a\ , \quad x_m > b \\ &\text{c)} \quad n < j < m \quad \text{implica} \quad a \leqslant x_j \leqslant b \end{aligned}$$

Por (3.5) , (3.3) e o fato de $\theta < a \leqslant x_j$

$$x_m - x_n < \frac{b-a}{3} + \sum_{j=n}^{m-1} a_j\ (\alpha - M(x_j)) < \frac{b-a}{3} + a_n\ (\alpha - M(x_n))$$

Se $x_n > \theta$, por (3.3) temos

$x_m - x_n < \dfrac{b-a}{3}$ que contradiz (3.6) - b)

Se $x_n < \theta$, por (3.1) temos

$$\text{(3.7)} \quad |M(x_n)| < c + d\ |x_n| < c + d\ |\theta| + + d\ |\theta - x_n| < c + d\ |\theta| + d\ (x_m - x_n)$$

Por (3.5)-b) e $\theta < a < x_j$ temos

$$\frac{x_m - x_n - (b-a)/3}{a_n} - |\alpha| < |M(x_n)|$$

Disto e de (3.7) temos

$$\frac{(x_m - x_n) - (b-a)/3}{a_n} - |\alpha| < c + d\,|\theta| + d(x_m - x_n)$$

Isto implica por (3.5)-a) que

$$(x_m - x_n)\,(1 - a_n\, d) < 2\,\frac{(b-a)}{3}$$

Que também por (3.5)-a) implica

$$x_m - x_n < b - a$$

Que contradiz (3.6) , provando o lema

Passemos agora ao teorema que garante a convergência de $\{X_n\}$ para θ com probabilidade um.

Teorema : Se as condições (3.1) , (3.2) , (3.3) e (3.4) estão satisfeitas, então $\{X_n\}$ converge para θ com probabilidade um.

Dem :

Suponhamos que $P\,\{\lim x_n = X\} = 1$, como exige o lema 3, e que

$$P\,\{X \neq \theta\} > 0$$

Consideremos ε_1 e ε_2 com $\theta < \varepsilon_1 < \varepsilon_2 < \infty$ tal que

$$P \{ \varepsilon_1 < X < \varepsilon_2 \} > 0$$

Caso isto não ocorra escolhemos ε_1 e ε_2 de modo que $-\infty < \varepsilon_1 < \varepsilon_2 < \theta$ e $P \{ \varepsilon_1 < X < \varepsilon_2 \} > 0$

Então para toda sequência $\{ x_n \}$ tal que $\lim x_n = x$ temos, para n suficientemente grande $\varepsilon_1 < x_n < \varepsilon_2$ e o conjunto de tais sequências tem probabilidade positiva.

Pelo lema 2 e lema 3 temos que para quase todas estas sequências

$$\Sigma a_i (\alpha - M(x_i)) \quad \text{converge}$$

Isto contradiz (3.4) e o fato de que $\sum_{n=1}^{\infty} a_n = \infty$ pois

$$\sum_{i=N}^{\infty} a_i (\alpha - M(x_i)) \geqslant \inf_{\varepsilon' < x_n < \varepsilon_2} \{ |M(x_i) - \alpha| \} \sum_{i=1}^{\infty} a_i = \infty$$

Enunciaremos agora os teoremas demonstrados por DVORETZKY [3] que engloba todos os resultados obtidos até aqui.

As demonstrações não serão colocadas pois as condições dos teoremas não nos parecem de fácil adaptação para as aplicações.

Teorema : Sejam $\{\alpha_n\}$, $\{\beta_n\}$ e $\{\Gamma_n\}$ sequências de números reais não negativas tal que

$$\lim_{n \to \infty} \alpha_n = 0$$

$$\sum_{n=1}^{\infty} \beta_n < \infty$$

$$\sum_{n=1}^{\infty} \Gamma_n = \infty$$

Seja θ um número real e T_n uma transformação mensurável satisfazendo

$$| T_n (x_1, \ldots , x_n) - \theta | < \max \{ \alpha_n, (1+\beta_n) (x_n - \theta) - \Gamma_n \}$$ para todo $x_1 , \ldots , x_n$

Sejam X_1 e Y_n $(n = 1,2, \ldots)$ variáveis aleatórias. Definimos

$$X_{n+1} = T_n (X_1,\ldots,X_n) + Y_n (X_1, \ldots, X_n) \qquad n > 1$$

Então as condições

$$E (X_1^2) < \infty \qquad \sum_{n=1}^{\infty} E (Y_n^2) < \infty$$

$E (Y_n / X_1, \ldots , X_n) = 0$ com probabilidade um para todo n, implicam

$$P \{ \lim_{n\to\infty} x_n = \theta \} = 1$$

Uma extensão do problema será :

Teorema : Sejam $\{ \alpha_n (x_1 , \ldots , x_n) \}$ e $\{ \beta_n (x_1, \ldots , x_n)\}$ sequências de funções não negativas tal que :

As funções $\alpha_n (x_1, \ldots , x_n)$ são uniformemente limitadas e $\lim_{n\to\infty} \alpha_n (x_1, \ldots , x_n) = 0$ uniformemente para toda sequência $x_1, \ldots , x_n, \ldots$

As funções $\beta_n (x_1 , \ldots , x_n)$ são mensuráveis e $\sum_{n=1}^{\infty} \beta_n (x_1, \ldots , x_n)$ é uniformemente limitada e uniformemente convergente para toda sequência $x_1 , \ldots , x_n , \ldots$

Para todo $L > 0$ existe funções $\Gamma_n (x_1,\ldots,x_n)$ satisfazendo

$$| T_n(x_1, \ldots , x_n) - \theta| < \max \{ \alpha_n ,(1 + \beta_n) |x_n- \theta| - \Gamma_n \}$$

onde T_n é uma transformação mensurável .

$\Sigma \Gamma_n (x_1 , \ldots , x_n) = \infty$ uniformemente para toda sequência $x_1 , \ldots , x_n , \ldots$ para a qual $\sup_{n=1,2,\ldots} | x_n | < L$

As demonstrações destes teoremas foram simplificadas por WOLFOWITZ | 12 | .

CAPÍTULO II

Localização do Máximo de uma Função de Regressão

O processo de ROBBINS - MONRO possibilitou a KIEFER - WOLFOWITZ [7] , com algumas modificações, estudar um método para a localização do máximo de uma função de regressão .

Desenvolveremos neste capítulo este método e a contribuição de BLUM [2] que mostra, com hipóteses mais fracas, a convergência com probabilidade um .

1 - *O Método de KIEFER - WOLFOWITZ*

Suponhamos que a cada valor real x , esteja associado uma variável aleatória $Y = Y(x)$ com distribuição $H(y/x)$ e média

$$M(x) = \int_{-\infty}^{\infty} y \, dH(y/x)$$

Sejam $\{a_n\}$ e $\{c_n\}$ sequências reais positivas

Admitimos que

(A) $M(x)$ é estritamente crescente para $x < \theta$ e estritamente decrescente para $x > \theta$

(B) $\int_{-\infty}^{\infty} (y - M(x))^2 \, dH(y/x) \leq \sigma^2$

(C) c_n converge para zero

$$\sum_{n=1}^{\infty} a_n = \infty$$

$$\sum_{n=1}^{\infty} a_n \ c_n < \infty$$

$$\sum_{n=1}^{\infty} a_n^2 \ c_n^2 < \infty$$

(D) Existem β e B positivos tal que

$$|x - \theta| + |x' - \theta| < \beta \quad \text{implica} \quad |M(x) - M(x')| < B \ |x-x'|$$

(E) Existem ρ e R positivos tal que

$$|x - x'| < \rho \qquad \text{implica} \qquad |M(x) - M(x')| < R$$

(F) Para todo δ , existe um $\Pi(\delta)$ positivo tal que

$$|\ x - \Theta\ | > \delta \qquad \text{implica} \qquad \inf_{o<\varepsilon<\frac{1}{2}\delta} \left| \frac{M(x+\varepsilon) - M(x-\varepsilon)}{\varepsilon} \right| > \Pi(\delta)$$

Seja x_1 um número arbitrário. Para todo n inteiro, definimos

$$X_{n+1} = X_n + \frac{a_n}{c_n} (Y_{2n} - Y_{2n-1})$$

onde Y_{2n} e Y_{2n-1} são variáveis aleatórias independentes com distribuições $H (y/x_n + c_n)$ e $H (y/x_n - c_n)$ respectivamente . Com as condições de regularidades exigidas acima, mostraremos que X_n converge estocasticamente para Θ .

Não exigimos a existência da derivada de $M(x)$ mas, a condição (D) diz que $M(x)$ é contínua numa vizinhança de Θ .

Se $M(x)$ tivesse derivada ela deveria ser zero para $x = \theta$, então nós esperamos que a derivada numa vizinhança de θ não seja muito grande. É o que pedimos em (D) .

Se a partir de uma certa distância de θ , $M(x)$ tivesse uma declividade muito pequena, o movimento na direção de θ seria muito lento. Fora de uma vizinhança de θ , gostaríamos que o módulo da declividade fosse limitada inferiormente por um número positivo, isto é o que pedimos em (F) .

Se $M(x)$ desse saltos muito bruscos em algumas partes , nós poderíamos por azar, dar um movimento para x que nos afastasse de θ . Se isto ocorresse em muitas partes, x poderia tender a $+\infty$ ou $-\infty$ com probabilidade positiva. Para que isto não ocorra, pedimos que os saltos de $M(x)$, caso existam, sejam limitados .

Na prática, se as condições impostas a $M(x)$ estiverem satisfeitas num intervalo $[C_1 , C_2]$ com $C_1 < \theta < C_2$ podemos aplicar o método para estimar θ . Supomos entretanto que $x_n \pm C_n$ caia fora do intervalo $[C_1 , C_2]$ e deste modo não podemos fazer uma observação neste ponto. Se deslocarmos x_n tal que $x_n \pm C_n$ atinja C_1 ou C_2 nós podemos fazer a observação e nossas conclusões continuam válidas .

Mostremos agora que X_n converge estocasticamente para θ .

Consideremos

(1.1) $b_n = E\{(X_n - \theta)^2\}$

(1.2) $U_n(x) = (x - \theta)\ E\ \{Y_{2n} - Y_{2n-1} \mid X_n = x\}$

(1.3) $U_n^+(x) = \frac{1}{2}(U_n(x) + |U_n(x)|)$,

$$U_n'(x) = \frac{1}{2}(U_n(x) - |U_n(x)|)$$

$$(1.4) \quad P_n = E\{U_n^+(x_n)\} \qquad N_n = E\{U_n^-(x_n)\}$$

$$(1.5) \quad e_n = E\{(Y_{2n} - Y_{2n-1})^2\}$$

Temos pela definição do processo

$$b_{n+1} = E\{(X_{n+1} - \Theta)^2\} = E\{|X_n - \Theta + \frac{a_n}{c_n}(Y_{2n} - Y_{2n-1})|^2\} =$$

$$= E\{(X_n - \Theta)^2\} + 2\frac{a_n}{c_n} E\{(X_n - \Theta)(Y_{2n} - Y_{2n-1})\} +$$

$$+ \frac{a_n^2}{c_n^2} E\{(Y_{2n} - Y_{2n-1})^2\} =$$

$$= b_n + 2\frac{a_n}{c_n} E\{(X_n - \Theta) E[Y_{2n} - Y_{2n-1} / x_n]\} + \frac{a_n^2}{c_n^2} e_n =$$

$$= b_n + 2\frac{a_n}{c_n}(P_n + N_n) + \frac{a_n^2}{c_n^2} e_n$$

Somando de um até n temos

$$(1.6) \quad b_{n+1} = b_1 + 2\sum_{i=1}^{n} \frac{a_i}{c_i} P_i + 2\sum_{i=1}^{n} \frac{a_i}{c_i} N_i +$$

$$+ \sum_{i=1}^{n} \frac{a_i^2}{c_i^2} e_i$$

Observamos que

$$U_n^+(x) \geq 0$$

$$U_n^+(x) > 0 \quad \text{implica} \quad |x - \Theta| < C_n$$

Mostremos esta implicação. Para isto, supomos

$$|x - \theta| \geqslant c_n$$

$$U_n(x) = (x - \Theta)\ E\ \{Y_{2n} - Y_{2n-1} / X_n = x\} =$$

$$= (x - \theta)\ \{M(x + c_n) - M(x - c_n)\}$$

Se $x \geqslant \theta$ então $x - \Theta \geqslant c_n$ e $x - c_n \geqslant \Theta$

Logo $x + c_n \geqslant \Theta$

Como $M(x)$ é decrescente para $x > \Theta$, $U_n(x) < 0$

Se $x < \theta$ então $x - \Theta \leqslant - c_n$ e $x + c_n \leqslant \Theta$

Logo $x - c_n \leqslant \Theta$

Como $M(x)$ é crescente para $x < \theta$, $U_n(x) < 0$

Sendo $U_n^+(x) > 0$ temos $U_n(x) = U_n^+(x) > 0$. Isto só ocorre se

$$|x - \Theta| < c_n$$

Para n suficientemente grande tal que $c_n < \dfrac{\beta}{4}$ temos

$$|x + c_n - \Theta| + |x - c_n - \Theta| < |x - \Theta| + c_n + |x - \Theta| + c_n <$$

$$< 4c_n < \beta \qquad \text{se} \quad U_n^+(x) > 0$$

Isto implica que

$$U_n^+(x) = (x - \Theta)\ |M(x + c_n) - M(x - c_n)| < c_n\ \beta\ 2c_n \quad \text{por (D)}$$

Temos, portanto

$0 \leq U_n^+(x) < 2 \; c_n^2$ mostrando assim que

(1.7) $\sum_{n=1}^{\infty} \frac{a_n}{c_n} P_n$ é convergente, digamos para α

Por (C) e (E) temos

$| M(x_n + c_n) - M(x_n - c_n) |^2 < R^2$ para n suficientemente grande

Então

(1.8) $E \{(Y_{2n} - Y_{2n-1})^2 / x_n \} = E \{(Y_{2n} - M(X_n+c_n))^2 +$

$+ (Y_{2n-1} - M(X_n - c_n))^2 / x_n \} + | M(X_n + c_n) - M(X_n - c_n)|^2 \leq$

$\leq 2 \sigma^2 + R^2$

Disto e de (C) obtemos que a série de termos positivos

(1.9) $\sum_{n=1}^{\infty} \frac{a_n^2}{c_n^2} e_n$ é convergente, digamos para Γ

Como $b_{n+1} \geq 0$ temos, por (1.6), (1.7) e (1.9),

$$2 \sum \frac{a_n}{c_n} N_n \geq - b_1 - 2\alpha - \Gamma > - \infty$$

Mostrando que a série de termos negativos

(1.10) $\sum \frac{a_n}{c_n} N_n$ é convergente .

Seja

$$k_n = \frac{| M(X_n + c_n) - M(X_n - c_n) |}{c_n}$$

Então

$$E \{ k_n \; |X_n - \Theta| \; \} = \frac{P_n - N_n}{c_n}$$

Logo

$$\lim \; \inf \; E \{ k_n \; | \; X_n - \Theta \; | \; \} = 0 \qquad \text{pois}$$

$$\Sigma \frac{a_n}{c_n} (P_n - N_n) \text{ é convergente por } (1.7) \text{ e } (1.10)$$

Seja $\{ n_i \}$ uma sequência de inteiros positivos tal que

$$\lim \; E \{ k_{n_i} \; | \; X_{n_i} - \Theta \; | \; \} = 0$$

Mostremos que $(X_{n_i} - \Theta)$ converge estocasticamente para zero.

Se isto não ocorrer, existirá δ e ε números positivos e uma subsequência $\{ t_i \}$ da sequência $\{ n_i \}$ tal que para todo t_i

$$P \; \{ \; | \; X_{t_i} - \Theta \; | \; > \delta \; \} > \varepsilon$$

Isto implica que

$$E \; \{k_{t_i} \; |X_{t_i} - \Theta| \} = E \; \frac{\{ |M(X_{t_i} + c_{t_i}) - M(X_{t_i} - c_{t_i}) \quad X_{t_i} | \}}{c_{t_i}} \geqslant$$

$$\geqslant \delta \; \varepsilon \; \pi(\frac{\delta}{2}) \text{ para todo } t_i \text{ tal que } c_{t_i} < \frac{1}{2} \; \delta$$

Mas isto contradiz o fato de que

$$\lim E \{k_{n_i} | X_{n_i} - \theta | \} = 0$$

Portanto

$(X_{n_i} - \theta)$ converge estocasticamente para zero .

Sejam η e ε números positivos arbitrários. Nós teremos demonstrado a convergência se existir um número $N(\eta, \varepsilon)$ tal que

$$P \{| X_n - \theta | > \eta \} < \varepsilon \qquad \text{para todo} \quad n > N(\eta, \varepsilon)$$

Seja s um número positivo tal que $\frac{s^2 + s}{2} < \varepsilon/2$

Como X_{n_i} converge estocasticamente para θ, existe um inteiro N_o tal que

$$P \{ |X_{N_o} - \theta| > s\} < \varepsilon/2$$

Podemos escolher N_o tal que

$$c_n < \min (\rho/2, \beta/2) \text{ para todo } n > N_o$$

$$\sum_{n=N_o}^{\infty} \frac{a_n^2}{c_n^2} \leqslant \frac{s^2}{2R^2 + 4^2}$$

$$\sum_{n=No}^{\infty} a_n c_n \leqslant \frac{s}{8B}$$

Observamos que

$$E \{ (X_{n+1} - \theta)^2 / X_{N_o} = x \} = E \{ E[(X_{n+1} - \theta)^2 / X_{N_o} = x, x_n] / X_{N_o} = x\} =$$

$$= E\{E[(X_n-\theta)^2/X_{N_o}=x, x_n]/X_{N_o}=x\} +$$

$$+ 2\frac{a_n}{c_n} E\{E[(x_n-\theta)(Y_{2n}-Y_{2n-1})/X_{N_o}=x, x_n]/X_{N_o}=x\} +$$

$$+\frac{a_n^2}{c_n^2} E\{E[(Y_{2n}-Y_{2n-1})^2/X_{N_o}=x, x_n]/X_{N_o}=x\} =$$

$$= E\{(X_n-\theta)^2/X_{N_o}=x\} + 2\frac{a_n}{c_n}E\{U_n(x_n)/X_{N_o}=x\} +$$

$$+\frac{a_n^2}{c_n^2} E\{E[(Y_{2n}-Y_{2n-1})^2/X_{N_o}=x, x_n]/X_{N_o}=x\}$$

Somando de N_o até n temos

$$b_{n+1}(x) = (x-\theta)^2 + 2\sum_{i=N_o}^{n}\frac{a_i}{c_i}E\{U_i(x_i)/X_{N_o}=x\} +$$

$$+\sum_{i=N_o}^{n}\frac{a_i^2}{c_i^2}E\{E[(Y_{2n}-Y_{2n-1})^2/X_{N_o}=x, x_n]/X_{N_o}=x\} \leqslant$$

$$\leqslant (x-\theta)^2 + 2\sum_{i=N_o}^{n}\frac{a_i}{c_i}E\{U_i^+(x_i)/X_{N_o}=x\} +$$

$$+\sum_{i=N_o}^{n}\frac{a_i^2}{c_i^2}(R^2+2\sigma^2) < (x-\theta)^2 + s$$

Por TCHEBYCHEFF temos

$$P\{|X_n-\theta|>\eta/|X_{N_o}-\theta|<s\} \leqslant \frac{E\{(X_n-\theta)^2/|X_{N_o}-\theta|<s\}}{\eta^2}$$

Mas

$$E\{(X_n-\theta)^2/|X_{N_o}-\theta|<s\} = \frac{1}{P\{|X_{N_o}-\theta|<s}\int_{|X_{N_o}-\theta|<s}(X_n-\theta)^2\,dP$$

$$= \frac{1}{P\{|X_{N_o} - \theta| < s\}} \int_{|x-\theta| < s} E\,(X_n - \theta)^2 / X_{N_o} = x\,\}\, dP(X_{N_o} \leqslant x) \leqslant$$

$$\leqslant \frac{1}{P\{|X_{N_o} - \theta| < s\}} \int_{|x-\theta| < s} \left[(x-\theta)^2 + s\right] dP\,(X_{N_o} \leqslant x) \leqslant$$

$$\leqslant \frac{1}{P\{|X_{N_o} - \theta| < s\}} \int_{|x-\theta| < s} (s^2 + s)\, dP\,(X_{N_o} \leqslant x) = s^2 + s$$

Portanto

$$P\{\,|X_n - \theta| > \eta / \,|X_{N_o} - \theta| < s\,\} \leqslant \frac{s^2+s}{\eta^2} < \varepsilon/2$$

Para $\eta > N_o$

$$P\{|X_n - \theta| > \eta\} = P\{\,|X_{N_o} - \theta| < s\,\}\; P\{|X_n - \theta| > \eta/\,|X_{N_o} - \theta| < s\} +$$

$$+ P\{\,|X_{N_o} - \theta| > s\,\}\; P\{\,|X_n - \theta| > \eta/\,|X_{N_o} - \theta| > s\,\} <$$

$$< P\{\,|X_{N_o} - \theta| < s\,\} \cdot \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\, P\{|X_n - \theta| > \eta \,|\, |X_{N_o} - \theta| > s\,\} <$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Mostrando que X_n converge para θ em probabilidade .

2 - *Convergência com Probabilidade Um*

O problema de localização de um máximo de uma função de regressão foi também estudado por BLUM [2] . Ele conseguiu, com hipóteses mais fracas que WOLFOWITZ , mostrar a convergência com probabilidade um da sequência $\{ X_n \}$ definida por

$$X_{n+1} = X_n + \frac{a_n}{c_n} (Y_{2n} - Y_{2n-1})$$

Faremos a demonstração deste fato usando, o lema 1 e com, algumas modificações, os lemas 2 e 3 e teorema 1 do capítulo I-3 .

Lema 1: (LOÈVE [8]) : Seja $\{ v_n \}$ uma sequência de variáveis aleatórias tal que $\Sigma E \{ v_n^2 \} < \infty$. Então $\Sigma \{ v_i - E (v_i / v_1 , \ldots , v_{i-1}) \}$ converge para uma variável aleatória com probabilidade um .

Consideremos as seguintes condições para $M(x)$ e as sequências $\{ a_n \}$ e $\{ c_n \}$

B. $\int_{-\infty}^{\infty} (Y - M(x))^2 \, dH (Y/x) \leqslant \sigma^2$

E. $M(x)$ crescente p/ $x < \Theta$

decrescente p/ $x > \Theta$

F. $\exists \rho , R$ tal que $|x' - x''| < \rho \Rightarrow |M(x') - M(x'')| < R$

G. $\forall \delta > o \quad \exists \Pi(\delta)$ tal que $| x - \Theta | > \delta \Rightarrow$

$$\inf_{\frac{1}{2}\delta > \varepsilon > o} \left\{ \frac{M(x+ \varepsilon) - M(x- \varepsilon)}{\varepsilon} \right\} > \Pi (\delta)$$

H. $c_n \to 0 \quad , \quad \Sigma a_n = \infty \quad , \quad \Sigma \left(\frac{a_n}{c_n}\right)^2 < \infty$

$$X_{n+1} = X_n + \frac{a_n}{c_n} (Y_{2n} - Y_{2n-1})$$

Lema 2 : Se B e H valem a sequência

$$X_{n+1} - \sum_{j=1}^{n} \frac{a_j}{c_j} \left(M(X_j + c_j) - M(x_j - c_j) \right)$$ converge com probabilidade um .

Dem.:

Seja $$v_j = \frac{a_j}{c_j} \left[Y_{2j} - M(X_j + c_j) - \right.$$
$$\left. - (Y_{2j-1} - M(X_j - c_j)) \right]$$

$$v_j^{\,2} = \frac{a_j^{\,2}}{c_j^{\,2}} \left[(Y_{2j} - M(X_j + c_j))^2 + (Y_{2j-1} - M(X_j - c_j))^2 - \right.$$
$$\left. - 2(Y_{2j} - M(X_j + c_j)) \cdot (Y_{2j-1} - M(X_j - c_j)) \right]$$

Como Y_{2j} e Y_{2j-1} são variáveis aleatórias independentes

$$E \{ (Y_{2j} - M(x_j + c_j)) (Y_{2j-1} - M(x_j - c_j)) \} =$$

$$= E (Y_{2j} - M(x_j + c_j)) \; E (Y_{2j-1} - M(x_j - c_j)) = 0$$

Logo

$$E \; v_j^{\,2} = \frac{a_j^{\,2}}{c_j^{\,2}} \left[E (Y_{2j} - M(x_j + c_j))^2 + \right.$$
$$\left. + E (Y_{2j-1} - M(x_j - c_j))^2 \right] \leqslant 2 \sigma^2 \frac{a_j^{\,2}}{c_j^{\,2}}$$

Por H

$$E \; v_j^{\,2} < \infty$$

Pelo lema 1 $\Sigma \{ v_j - E(v_j / v_1 , \ldots , v_{j-1}) \}$ converge

Dado x_1 temos $M(X_1 + c_1)$, $M(X_1 - c_1)$. Mas dado

$Y_2 - M(X_1 + c_1) - (Y_1 - M(X_1 - c_1))$ e $M(X_1 + c_1)$, $M(X_1 - c_1)$

nós temos X_2 , etc ...

Deste modo obtemos

$$E \{(Y_n - M(X_n + c_n) - (Y_{2n-1} - M(X_n - c_n)) \mid Y_2 - M(X_1 + c_1) -$$

$$- (Y_1 - M(X_1 - c_1)) , \ldots , Y_{2(n-1)} - M(X_{n-1} + c_{n-1}) -$$

$$- (Y_{2(n-1)-1} - M(X_{n-1} - c_{n-1}) \} = E \{ Y_{2n} - M(X_n + c_n) -$$

$$- (Y_{2n-1} - M(x_n - c_n)) / x_n \} = 0$$

Portanto

$$\Sigma \frac{a_j}{c_j} [Y_{2j} - M(X_j + c_j) - (Y_{2j-1} - M(X_j - c_j))] \quad \text{converge}$$

com probabilidade um .

Como

$$X_{n+1} - \sum^{n} \frac{a_j}{c_j} [M(X_j + c_j) - M(X_j - c_j)] =$$

$$= X_1 + \sum^{n} \frac{a_j}{c_j} [Y_{2j} - Y_{2j-1}] + \sum^{n} \frac{a_j}{c_j} [- M(X_j + c_j) +$$

$$+ M(X_j - c_j)] = X_1 + \sum^{n} \frac{a_j}{c_j} [Y_{2j} - M(X_j + c_j) -$$

$$- (Y_{21-1} - M(X_j - c_j))]$$

temos o lema demonstrado .

Lema 3 : Se B. E. F. H. valem então X_n converge com probabilidade um.

Dem.:

Para demonstrar o lema, observamos que

$$(2.1) \quad P \{ \lim x_n = \infty \} + P \{ \lim x_n = -\infty \} = 0 \quad \text{pois}$$

Se $\lim x_n = \infty$ então $x_n > \theta$ a partir de um certo N_o

logo p/ $n > N_o$

$M(x_n + c_n) - M(x_n - c_n) < 0$ pois M é decrescente p/ $X > \theta$

Portanto

$$\lim x_{n+1} - \sum_{j=1}^{n} \frac{a_j}{c_j} [M(x_j + c_j) - M(x_j - c_j)] = \infty$$

mas isto acontece somente com probabilidade zero pelo Lema 2

Suponhamos que o lema seja falso e em vista de (2.1) existirá um conjunto amostral com probabilidade positiva tal que

(22)-a) $x_{n+1} - \sum \frac{a_j}{c_j} (M(x_j + c_j) - M(x_j - c_j))$ é convergente

(22)-b) $\quad \liminf x_n < \limsup x_n$

para toda sequência neste conjunto.

Seja $\{X_n\}$ uma tal sequência e suponhamos que $\limsup x_n > \theta$

(Argumento semelhante se $\limsup X_n < \theta$)

Sejam a e b satisfazendo

$$a > \theta$$

$$\liminf X_n < a < b < \limsup X_n$$

Seja N tal que $m, n > N$

$$(23) \quad \left| X_m - X_n - \sum_{j=n}^{m-1} \frac{a_j}{c_j} M(X_j + c_j) - M(X_j - c_j) \right| < \frac{b-a}{3}$$

que é possivel por (22)-a)

Escolhemos m e n tal que

(24) $\quad N < n < m$

(25) $\quad X_n < a \quad , \quad X_m > b$

(26) $\quad n < j < m \implies a \leqslant x_j \leqslant b$

(27) $\quad R \dfrac{a_n}{c_n} < \dfrac{b-a}{3}$

(28) $\quad 2\, c_n < \rho$

Por (23) e (26)

$$(29) \quad X_m - X_n < \frac{b-a}{3} + \sum_{j=n}^{m-1} \frac{a_j}{c_j} \left(M(X_j + c_j) - M(X_j - c_j) \right) \leqslant$$

$$\leq \frac{b-a}{3} + \frac{a_n}{c_n} \left(M(X_n + c_n) - M(X_n - c_n) \right)$$

Se $X_n > \theta$ por (29) e E .

$$X_m - X_n < \frac{b-a}{3} \quad \text{que contradiz (25)}$$

Se $X_n < \theta$, sendo $2c_n < R$ temos

$$| X_n + c_n - X_n + c_n | = 2c_n < \rho \Rightarrow |M(X_n + c_n) - M(X_n - c_n)| < R$$

Então

$$X_m - X_n \leq \frac{b-a}{3} + R \frac{a_n}{b_n} < \frac{b-a}{3} + \frac{b-a}{3} = \frac{2.(b-a)}{3}$$

que contradiz (25).

Teorema : Se B. E. F. G. H. valem, então $X_n \to \theta$ com probabilidade um .

Dem.:

Suponhamos que $P \{ \lim X_n = X \} = 1$, como requer o lema 3, e que

$$P \{ X \neq \theta \} > 0$$

Sejam ε_1 e ε_2 tal que

$$\theta < \varepsilon_1 < \varepsilon_2 < \infty$$

$P \{ \varepsilon_1 < x < \varepsilon_2 \} > 0$ (caso contrário, $-\infty < \varepsilon_1 < \varepsilon_2 < \theta$

tal que

$$P\ \{\varepsilon_1 < x < \varepsilon_2\} > 0$$

Então para toda sequência amostral tal que $\lim x_n = x$ com $\varepsilon_1 < x < \varepsilon_2$ nós temos $\varepsilon_1 < x_n < \varepsilon_2$ p/ n suficientemente grande .

O conjunto de sequências amostrais que satisfazem esta propriedade tem probabilidade positiva .

Pelo lema 2 e lema 3 quase todas estas sequências são tais que

$$\Sigma\ \frac{a_j}{c_j}\left[\,M(x_j + c_j) - M(x_j - c_j)\,\right] \quad \text{é convergente}$$

Mas por G sendo $|\theta - X_n| > \varepsilon_1$

$$\Sigma\ \frac{a_j}{c_j}\left[M(X_j + c_j) - M(X_j - c_j)\right] > \Sigma\ a_j\ \Pi(\varepsilon_1) = \infty$$ que

contradiz o fato de

$$\Sigma\ \frac{a_j}{c_j}\left[M(X_j + c_j) - M(X_j - c_j)\right] < \infty$$ provando o

Teorema .

CAPÍTULO III

Aplicações

O objetivo do trabalho foi colocar o problema de aproximação estocástica de modo que as hipóteses sejam simples e claras, para facilitar as aplicações, não perdendo a generalidade.

Descreveremos neste capítulo o trabalho de GUTTMAN - GUTTMAN [5] que faz uma aplicação do método para um problema em biologia. Em seguida simularemos experimentos e compararemos o valor estimado com o método de aproximação estocástica e o valor real.

1. *Uma Aplicação em Biologia*

Deixando o paramécium caudatun sob o efeito da cinetina, obteremos uma multiplicação de células maior do que a não exposta. Esta é uma função crescente com o tempo. Com esta observação GUTTMAN - GUTTMAN [5] apresentam um trabalho que consiste em determinar um tempo Θ em que haja, em média, uma multiplicação α de células.

Tomamos um controle e o paramécium sujeito a cinetina. Após um certo número de horas, contamos as células do paramécium e determinamos a razão K/c, isto é, o número de células do sujeito a cinetina sobre o número de células do não sujeito.

O problema será determinar um tempo Θ de modo que em média temos $K/c = 1.1$.

Podemos usar o método de ROBBINS - MONRO para estimar θ . Para isto consideramos :

x_n = número de horas de tratamento no n-ésimo ensaio

y_n = K/c obtido após um ensaio de x_n horas

α = 1.1

θ = número de horas de tratamento que, em média, obteremos uma razão α

a_n é uma sequência satisfazendo $\Sigma a_n = \infty$, $\Sigma a_n^2 < \infty$. Neste experimento tomaremos $a_n = \frac{20}{n}$. Mais adiante daremos os motivos para esta escolha.

Para uma primeira aproximação, foi tomado x , igual a 30 horas. No primeiro ensaio as células foram contadas após 30 horas. Computamos então K/c e achamos 1.067 (veja tabela 1) . Substituindo em

$$x_{n+1} = x_n + a_n (\alpha - y_n)$$

achamos

$$x_2 = 30 + 20(1.1 - 1.067) = 30.66 \text{ horas}$$

No segundo ensaio, as células foram contadas após 30 horas e 40 minutos. A razão obtida para este tempo foi $\frac{K}{c} = 1.30$, isto é, $y_2 = 1.30$. Obtemos então x_3 por

$$x_3 = 30.7 + \frac{20}{2} (1.1 - 1.3) = 28.7 \text{ horas}$$

Para cada estágio da experimentação, um novo tempo de tratamento foi tomado de acordo com o resultado do ensaio anterior, o α proposto e o número de ensaios já feitos.

Notamos na Tabela 1 que a partir do sexto ensaio, x_n começa a estabilizar em torno de 25 horas.

TABELA 1

N	x_n	y_n	x_{n+1}
1	30	1.067	30.7
2	30.7	1.30	28.7
3	28.7	1.131	27.3
4	27.3	1.223	26.6
5	26.6	1.577	24.8
6	24.8	1.133	24.6
7	24.6	0.89	25.2
8	25.2	1.00	25.5
9	25.5	0.81	25.6
10	25.6	1.31	25.1
11	25.1	1.21	24.8
12	24.8	1.03	24.9

Tínhamos por hipótese que o número de 30 horas, levaria a um resultado menor que 1.5 e maior que 1.0 . Esperávamos que este tempo fosse grande mas que não ultrapassasse mais que 10 horas do tempo procurado. A diferença do observado para 1.1 seria no máximo de 0.4 . Escolhemos então 20 que nos daria a correção 20 x 0.4 = 8 horas. Deste modo $a_n = \frac{20}{n}$ parecia ser uma boa sequência.

Notamos na Tabela 1 que paramos no 13º ensaio. O motivo para isto foi que a partir do 6º ensaio, o número de horas ficou em torno de 25 . A variação não era muito gran e para o problema. Deste modo tomamos $\Theta = 25$ horas.

O mesmo experimento foi montado tomando $\alpha = 1.05$.

Pelo problema anterior, aceitamos que para 25 horas tínhamos em média a razão K/c igual a 1.1 . Decidimos então que o primeiro valor x , seria de 23 horas. Sabíamos que este valor observado não deveria ultrapassar 1.1 . Teríamos deste modo uma diferença, do real para o observado, de no máximo 0.05 . A sequência $\{ a_n \}$ foi tomada da forma $\frac{30}{n}$ pois no primeiro passo teríamos uma correção de $30 \times 0.05 = 1.5$ horas. Obtivemos então os seguintes valores :

TABELA 2 .

N	x_n	y_n	x_{n+1}
1	23	1.14	20.3
2	20.3	1.00	21.1
3	21.1	1.75	14.1
4	14.1	0.97	14.7
5	14.7	1.26	13.4
6	13.4	0.98	13.8
7	13.8	1.04	13.8
8	13.8	0.96	13.5
9	13.5	1.25	12.8
10	12.8	1.22	12.3
11	12.3	1.04	12.3
12	12.3	1.11	12.2
13	12.2	1.17	11.8
14	11.8	0.97	12.0
15	12.0	0.81	12.5

Observamos que após o nono ensaio, x_n começa a estabilizar e o último tempo determinado é de 12.5 horas. Aceitamos então 12.5 horas como o tempo para que tenhamos em média $\frac{K}{c} = 1.05$.

É curioso notar que a média das observações (K/c) dos últimos seis experimentos é 1.05 . Isto parece uma confirmação de que o tempo deverá ser 12.5 horas .

2. *Um Modelo Teórico*

Para percebermos melhor a importancia da escolha da sequência $\{ a_n \}$ e também para termos uma idéia dos erros cometidos, criamos um experimento onde seja possível a aplicação do processo de ROBBINS - MONRO .

Conhecemos a priori a distribuição de $Y(x)$ bem como a função $M(x) = E(Y(x))$. Deste modo conhecemos a raiz Θ da equação $M(x) = \alpha$. Aplicaremos o processo de ROBBINS - MONRO para estimar Θ e depois discutiremos a possibilidade de aplicação deste processo e compararemos o valor real Θ com o valor estimado .

Suponhamos um experimento onde a cada número real x , esteja associado uma variável aleatória $Y(x)$. Queremos achar a raiz Θ da equação $M(x) = 3$. Sabemos que é possível a aplicação do processo de ROBBINS - MONRO e que o intervalo onde $M(x)$ satisfaz as hipóteses necessárias para a aplicação é $[-9,9]$. Como mais nada foi dito sobre $M(x)$, usaremos, numa primeira vez, a sequência $a_n = \frac{1}{n}$ e o valor inicial $x_1 = 2$.

Com o valor $x_1 = 2$ conduzimos o experimento e obser vamos $y_1 = 7.46$ da variável aleatória $Y(2)$. Determinamos x_2 por

$$x_2 = 2 + 1(3 - 7.46) = -2.46$$

Com este valor $x_2 = -2.46$ observamos $y_2 = -17.703$ da variável aleatória $Y(-2.46)$. Calculamos x_3 por

$$x_3 = -2.46 + \frac{1}{2}(3 - (-17.703)) = 7.891$$

Para x_4 obtivemos -155.062 (veja Tabela 3). Este ponto está fora do intervalo $[-9,9]$ e não podemos fazer observação de $Y(-155.062)$ pois as hipóteses sobre $M(x)$ não estão satisfeitas fora deste intervalo. Para uma nova observação, usamos $x_4 = -9$ para que possamos continuar com o experimento.

Deste modo fizemos quarenta observações. Ora tomamos 9 ora -9 para x_n. Obtivemos $x_{41} = -8.675$ que pertence ao intervalo $[-9,9]$ e a partir daí fomos tomando os x_n por

$$x_{n+1} = x_n + \frac{1}{n}(3 - y_n) \quad \text{até} \quad n = 70$$

TABELA 3

N	x_n	y_n	x_{n+1}
1	2.000	7.460	- 2.460
2	-2.460	- 17.730	7.891
3	7.891	491.861	-155.062
4	-9.0	-727.655	173.664
5	9.0	729.794	-136.359
6	-9.0	-731.111	113.352

cont.

continuação :

N	x_n	y_n	x_{n+1}
7	9.0	727.631	- 94.519
⋮	⋮	⋮	⋮
20	-9	-727.573	27.529
⋮	⋮	⋮	⋮
30	-9	-728.788	15.393
⋮	⋮	⋮	⋮
40	-9	-724.603	9.190
41	9	727.676	- 8.675
42	-8.675	-651.195	6.901
43	6.901	328.679	- 0.673
44	-0.673	2.753	- 0.667
45	-0.667	1.122	- 0.626
46	-0.626	1.521	- 0.593
47	-0.593	- 1.012	- 0.508
⋮	⋮	⋮	⋮
67	0.435	3.769	0.424
68	0.424	1.032	0.453
69	0.453	- 0.487	0.503
70	0.503	- 0.499	0.553

Vemos que para $n = 43$, $x_{n+1} = -0.673$ e que a partir daí x_n cresce lentamente até a última estimativa que foi para $n = 70$. Então, com estas observações, nada podemos dizer sobre Θ pois x_n não se estabilizou.

Isto ocorreu pelo fato de não termos feito nenhum estudo sobre $M(x)$ e não termos tomado uma sequência $\{ a_n \}$ mais adequada ao problema.

Tendo em vista esta primeira tentativa para estimar Θ tomaremos $a_n = 0.05/n$. Esta sequência foi tomada pois as correções não deveriam ser maiores que nove que é o raio do intervalo. Temos então no terceiro ensaio, uma correção em x_2 de

$$\frac{0.05}{3} \; (-3 + 491.861) = 8.148$$

Com esta sequência $\{ a_n\}$ repetimos o experimento e obtivemos a Tabela 4 . Vemos que após o quinto ensaio x_n estabiliza em torno de 1.4. O menor valor achado foi $x_n = 1.428$ ($n = 10$) e o maior foi $x_n = 1.492$ ($n = 6$) . Podemos dizer então que $\Theta = 1.4$. Como a partir do décimo sétimo ensaio x_n estabiliza-se em torno de 1.46, isto poderia nos levar a dizer que $\Theta = 1.46$.

Para este experimento, tomamos $Y(x)$ uma variável aleatória tendo distribuição normal com média x^3 e variância dois . Estávamos estimando a solução de $E (Y(x)) = 3$ que seria então 1.442 pois $(1.442)^3 = 3$. Com o processo de ROBBINS - MONRO, obtivemos o ponto 1.4 ou 1.46 que é bastante próximo do real .

TABELA 4

N	x_n	y_n	x_{n+1}
1	2	9.904	1.655
2	1.655	8.075	1.528
3	1.528	2.898	1.530
4	1.530	4.582	1.510
5	1.510	4.803	1.492
6	1.492	5.410	1.472
7	1.472	4.626	1.460
8	1.460	4.909	1.448
9	1.448	6.574	1.428
10	1.428	1.612	1.435
11	1.435	3.474	1.433
12	1.433	0.847	1.442
13	1.442	0.704	1.451
14	1.451	4.387	1.446
15	1.446	2.285	1.448
16	1.448	1.336	1.454
17	1.454	0.005	1.462
18	1.462	3.642	1.461
19	1.461	2.855	1.461
20	1.461	0.909	1.466
21	1.466	2.003	1.469
22	1.469	4.442	1.465
23	1.465	3.119	1.465
24	1.465	0.257	1.471
25	1.471	5.122	1.466

cont.

continuação :

N	x_n	y_n	x_{n+1}
26	1.466	2.004	1.468
27	1.468	2.213	1.470
28	1.470	5.388	1.466
29	1.466	2.579	1.466
30	1.466	-0.474	1.472

A possibilidade de aplicação deste método é que a função $M(x) = x^3$ e a familia $Y(x)_{\, x \,\varepsilon\, R}$ satisfazem :

(1) $|M(x)| = |x^3| < 81 \; |x| \quad p/ \; x \;\varepsilon\; [-9,9]$

(2) $\int_{-\infty}^{\infty} | Y - M(x)|^2 \;\alpha\; H(Y/x) = E \; | Y(x) - E(Y(x))|^2 =$
$= \text{var } Y(x) = 2 < \infty$

(3) $M(x) < 3 \quad p/ \; x < \Theta \; , \quad M(x) > 3 \quad p/ \; x > \Theta$
pois $M(x)$ é crescente

(4) $\inf\limits_{\delta_1 < |x - \Theta| < \delta_2} \{|M(x) - 2|\} > 0$ pois $M(x)$ é crescente

Estas são as condições (3.1), (3.2), (3.3) e (3.4) exigidas no capítulo I-3 que garantem a convergência com probabilidade um.

O mesmo experimento foi montado, mas agora tomamos $\alpha = 8$ $a_n = \frac{0.3}{n}$ e $x_1 = 3$. Notamos que no último ensaio obti vemos $x_{n+1} = 2$. Isto não quer dizer que obrigatoriamente $\theta = 2$. Os resultados estão na Tabela 5 .

TABELA 5

N	x_n	y_n	x_{n+1}
1	3	26.850	-2.66
2	-2.66	-16.157	0.969
3	0.969	- 3.678	2.136
4	2.136	10.643	1.939
5	1.939	2.307	2.280
6	2.280	10.111	2.175
7	2.175	10.270	2.078
8	2.078	10.254	1.993
9	1.993	5.754	2.068
10	2.068	9.968	2.009
11	2.009	10.730	1.934
12	1.934	9.590	1.895
13	1.895	9.615	1.857
14	1.857	6.154	1.897
15	1.897	8.942	1.878
16	1.878	6.989	1.897
17	1.897	10.128	1.860
18	1.860	8.803	1.846
19	1.846	4.223	1.906
20	1.906	7.301	1.916
21	1.916	10.703	1.878
22	1.878	3.083	1.945
23	1.945	3.418	2.004
24	2.004	8.622	1.997
25	1.997	7.748	2.000

3. *Máximo de uma Função de Regressão*

Como no exemplo anterior, para a aplicação do processo de ROBBINS - MONRO, criamos um modelo teórico onde seja possível a aplicação do processo de KIEFER - WOLFOWITZ para estimar o máximo de $E(Y(x))$.

Para isto tomamos a familia $\{ Y(x) \quad x \in R \}$ de variáveis aleatórias tendo, cada uma, distribuição normal com média $-x^2 + 2x$ e variancia um. Nosso experimento consistirá em tirar amostra de $Y(x)$ que faremos com uma tabela de números aleatórios normais.

Desejamos estimar o ponto de máximo de $M(x)$. Vemos facilmente, que esta função satisfaz as hipóteses do capítulo II-2 para $x \in (-3.3)$ e portanto, podemos aplicar o processo de KIEFER - WOLFOWITZ.

Na aplicação do processo, usamos as sequências $a_n = n^{-1}$ e $C_n = n^{-1/3}$. O primeiro valor proposto foi $x_1 = 0$ (veja Tabela 6). Analisando os resultados obtidos, vemos que existe uma variação muito grande. Mesmo após 25 ensaios temos o menor valor igual a 0.62 e o maior igual a 1.45 . Isto ocorre por não tomarmos as sequências $\{a_n\}$ e $\{c_n\}$ mais adequada ao problema.

TABELA 6

N	x_n	y_{2n}	y_{2n-1}	x_{n+1}
1	0	-0.87	-2.97	2.10
2	2.10	-2.03	0.88	0.27
3	0.27	1.27	-3.18	2.41
4	2.41	-4.39	-0.10	0.70
5	0.70	3.12	0.78	1.50
6	-1.50	-0.34	1.65	0.90
7	0.9	0.44	1.71	0.55
8	0.55	1.25	-0.88	1.08
9	1.08	0.77	1.71	0.87
10	0.87	0.62	1.69	0.64
11	0.64	-1.10	-0.10	0.43
12	0.43	2.84	-0.75	1.12
13	1.12	1.00	-0.96	1.47
14	1.47	-0.39	1.15	1.21
15	1.21	-0.33	1.30	0.94
16	0.94	0.04	0.16	0.92
17	0.92	2.00	0.37	1.17
18	1.17	0.46	2.74	0.84
19	0.84	-0.10	1.40	0.63
20	0.63	0.78	-1.86	0.98
21	0.98	-0.65	0.51	0.83
22	0.83	-0.07	0.45	0.76
23	0.76	3.00	0.10	1.12
24	1.12	1.00	1.13	1.09
25	1.09	1.61	0.57	1.21
26	1.21	1.62	0.24	1.37
27	1.37	1.76	1.07	1.45
28	1.45	0.37	1.09	1.37
29	1.37	1.67	2.80	1.25
30	1.25	0.97	0.63	1.28
31	1.28	2.18	1.28	1.37
32	1.37	0.51	2.05	1.22
33	1.22	1.26	1.23	1.22
34	1.22	0.75	2.64	1.04
35	1.04	-0.83	0.37	0.93

cont.

continuação :

N	x_n	y_{2n}	y_{2n-1}	x_{n+1}
36	0.93	1.48	1.38	0.94
37	0.94	1.05	0.22	1.02
38	1.02	1.58	1.11	1.06
39	1.06	0.60	1.96	0.94
40	0.94	0.43	0.10	0.97
41	0.97	0.08	2.04	0.77
42	0.77	2.88	2.92	0.77
43	0.77	1.66	-0.38	0.94
44	0.94	0.13	1.03	0.86
45	0.86	0.23	1.72	0.75
46	0.75	0.42	2.05	0.62
47	0.62	1.00	-0.17	0.71
48	0.71	1.55	-0.34	0.85
49	0.85	3.18	0.38	1.06
50	1.06	1.34	1.33	1.06

O mesmo experimento foi montado mas com $a_n = \frac{n^{-1}}{5}$, $c_n = n^{-1/3}$ e o mesmo valor inicial $x_1 = 0$ (veja Tabela 7) Analisando a Tabela 7 , vemos que x_n estabiliza-se rapidamente em torno de 0.9 . Poderiãmos então dizer que $\Theta = 0.9$.

Como $M(x) = -x^2 + 2x$, sabemos que o ponto de máximo desta função é um. Comparando este valor com o estimado, vemos que a diferença é pequena.

TABELA 7

N	x_n	y_{2n}	y_{2n-1}	x_{n+1}
1	0	-0.37	-4.10	0.75
2	0.75	1.69	0.55	0.89
3	0.89	0.97	-0.98	1.08
4	1.08	0.64	1.24	1.03
5	1.03	1.65	2.09	1.00
6	1.00	0.83	0.34	1.03
7	1.03	0.65	1.49	0.98
8	0.98	-0.28	1.66	0.88
9	0.88	1.27	1.05	0.89
10	0.89	1.64	0.60	0.94
11	0.94	1.48	1.12	0.95
12	0.95	2.05	1.25	0.98
13	0.98	-0.04	0.21	0.98
14	0.98	0.77	2.88	0.90
15	0.90	1.21	0.96	0.91
16	0.91	1.20	1.15	0.88
17	0.88	1.40	0.77	0.91
18	0.91	0.88	-0.63	0.95
19	0.95	0.10	1.57	0.91
20	0.91	-0.55	0.24	0.89

4. *Comentários sobre os Processos*

Nos processos de aproximação de ROBBINS - MONRO e KIEFER - WOLFOWITZ notamos um inconveniente na aplicação que é o fato de não termos uma regra de parada. Isto nos impossibilita afirmar qual o erro que cometemos quando paramos o processo .

Estudo sobre este fato foi feito por BLOCK [1] para o caso de ROBBINS - MONRO, mas exige o conhecimento de $E\ (X_1 - \theta)^2 = V^2$ que não pode ser usado pois não conhecemos

θ . Neste trabalho é mostrado qual a melhor sequência $\{ a_n \}$ e qual o erro quadrático médio $(E \{ (X_n - \theta)^2 \})$ que cometemos quando paramos o processo. A sequência $\{ a_n \}$ e a limitação são funções de $V^2 = E (X_1 - \theta)^2$.

Vimos na prática que este problema pode ser resolvido em parte, tendo alguns conhecimentos superficiais do que queremos estimar e algumas observações do experimento .

Uma facilidade que este processo traz é o fato de não exigir o conhecimento da distribuição de $Y(x)$. Nos modelos teóricos apresentamos a distribuição apenas para que pudessemos comparar o estimado com o real.

Na aplicação do processo, para que consigamos uma convergência mais rápida, podemos mudar o valor de x_n em qualquer ponto, pois o processo não depende do ponto inicial. Por exemplo:

1) Na Tabela 3 , vemos que para $x_4 = -9$ o valor observado y_4 era -727.655 que estava muito longe de $\alpha = 3$. Deveríamos neste ponto tomar x_5 bem menor que 9 pois para este valor observamos $y_5 = 729.794$. Com esta modificação aproximaríamos de θ mais rapidamente .

2) Na mesma Tabela 3 a partir de $n = 43$, vemos que x_n cresce lentamente. Poderíamos tomar, por exemplo, $x_{47} = 0$. Se x_n continuasse a crescer lentamente, poderíamos em qualquer ponto dar uma "ajuda" ao processo para que a convergência fosse mais rápida.

Sempre que estivermos adaptando a sequência $\{ a_n \}$ ou dando uma "ajuda" ao processo, devemos estar com a atenção voltada para o experimento e não somente aos números que obtemos.

BIBLIOGRAFIA

[1] BLOCK, H. D. Estimates of error for two modifications of the Robbins-Monro stochastic approximation process. Ann. Math, Stat. 28 (1957) 1003 - 1010.

[2] BLUM, J. R. Aproximation methods which converge with probality one. Ann. Math. Stat. 25 (1954) 382-386 .

[3] DVORETZKY, A. On stochastic approximation. Proceedings of the third Berkeley Symposium on Mathematical Statistic on Probability (1956) 39-55 .

[4] FRIEDMAN, M. and SAVAGE, L.J. Experimental determination of the maximum of a function. Selected techniques of statistical Analysis. McGraw-Hill, New York, (1947) 363 - 372 .

[5] GUTTMAN, L. and GUTTMAN, R. An ilustration of the use of stochastic approximation. Biometrics, 15 (1959) 551 - 559 .

[6] HOTELLING, H. Experimental determination of the maximum of a function. Ann. Math. Stat. 12 (1941) 20 - 46

[7] KIEFER, J. and WOLFOWITZ, J. Stochastic estimation of the maximum of a regression function. Ann. Math. Stat. 23 (1952) 462 - 466 .

[8] LOÈVE, M. On almost sure converge. Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability. (1951) 279 - 303 .

[9] ROBBINS. H. and MONRO, S. A stochastic approximation method. Ann. Math. Stat. 22 (1951) 400 - 407 .

[10] WASAN, M. T. Stochastic Approximation. Cambridge University Press. New York (1969) .

[11] WOLFOWITZ, J. On the stochastic approximation method of Robbins and Monro. Ann. Math. Stat. 23 (1952) 457 - 461 .

[12] WOLFOWITZ, J. On stochastic approximation method . Ann. Math. Stat. 27 (1956) 1151 - 1156 .