

# **Estimação da média com observações em falta**

Adilson de Jesus Martins da Silva





# Estimação da média com observações em falta

Adilson de Jesus Martins da Silva

Dissertação para a obtenção do Grau de **Mestre em Matemática**  
Área de Especialização em **Estatística, Optimização e Matemática financeira**

## Júri

**Presidente:** Carlos Manuel Rebelo Tenreiro da Cruz

**Orientador:** Paulo Eduardo Aragão Aleixo Neves de Oliveira

**Vogal:** Ana Cristina Martins Rosa

**Data:** Junho de 2010



# Resumo

Neste trabalho abordamos o problema da estimação da média de uma variável aleatória que depende de outra (ou outras) através de um modelo de regressão. O conhecimento completo da amostra permite o tratamento deste problema utilizando os métodos clássicos para a estimação da regressão. Aqui iremos estudar uma solução para o problema quando parte da informação que compõe a amostra se perde, nomeadamente, quando, na amostra que permite o tratamento do problema de regressão, algumas das ordenadas não são conhecidas, mas sim apenas o valor para variável explicativa. A solução mais simples neste caso consistiria em ignorar os pontos correspondentes da amostra. Contudo, aqui, pretende-se estudar um método que tenta tirar partido do conhecimento de que houve uma observação e que o valor desta está relacionado com a abcissa disponível através de um modelo de regressão que será aproximado à custa dos pontos da amostra conhecidos na totalidade.

Numa primeira instância introduzimos os métodos não paramétricos de aproximação da regressão, a explorar, regressograma e Nadaraya-Watson. Em capítulos posteriores, descrevemos a adaptação do estimador Nadaraya-Watson ao problema das amostras incompletas. Discutem-se ainda adaptações que tentam minimizar as dificuldades deste estimador quando o espaço das abcissas é de dimensão grande, assumindo que parte da função de regressão tem um comportamento linear, portanto paramétrico. Neste trabalho apresentamos resultados que asseguram a normalidade assintótica dos estimadores definidos.

**Palavras Chave:** Função de regressão, Regressograma, Nadaraya-Watson, Estimação da média, Observações em falta.

# Abstract

This work presents an approach for the estimation of the mean of a random variables that is influenced by other variables through a regression model. The complete knowledge of a sample enables treating this problem using classical methods for regression estimation. Here we will be interested in describing an approach when part of the sample has been lost, namely, when the value of the depend variable is not available. The simplest solution would be to ignore the points in the sample with incomplete information. We describe an approach that uses the regression model to infer the values that were lost given that we know the abscissae were the observation had been made.

We will first introduce the nonparametric methods to approach the regression function: regressogram and Nadaraya-Watson. The discussion follows by adapting the Nadaraya-Watson estimator to the incomplete

sample problem. We also discuss one way to go around the difficulties of this estimator when the abscissae space has a large dimension, assuming it is reasonable to accept that part of the regression function has a linear behaviour, turning this part into a parametric problem. We present results that prove the asymptotic normality of the above mentioned estimators.

**Keywords:** Regression function, Regressogram, Nadaraya-Watson, Mean estimation, Missing observation.







# Agradecimentos

*É com muita satisfação que, neste pequeno texto, registo o mais profundo agradecimento àqueles que de várias formas contribuíram para que a realização deste trabalho fosse possível. Sendo o espaço destinado para o efeito relativamente pequeno, não me permitindo citar os nomes de todos, resta-me, desde já, endereçar os meus mais sinceros agradecimentos àqueles que directa ou indirectamente me apoiaram durante todo o percurso académico, cujos nomes não figuram neste texto.*

*Agradeço:*

*Ao meu orientador o Professor Doutor Paulo Eduardo A. A. Neves de Oliveira a quem dedico este trabalho, por ter-me acolhido como orientando, pelas inúmeras horas dispensadas em virtude do esclarecimento das minhas incessantes dúvidas que, num clima descontraído, permitiu que a sua inestimável destreza matemática fluísse, estimulando em mim a paixão pela investigação e cultivando o meu orgulho pela área que, desde a minha infância, despertou o meu interesse, pela simpatia gratuita e pela dedicação incondicional para que este trabalho não fique longe do conjunto dos aplaudidos.*

*Ao Professor Doutor Pedro H. F. Q. de Almeida, pela simpatia e prontidão em facultar materiais, cuja falta dificultaria a elaboração deste trabalho.*

*Aos professores doutores, Alexander V. M. Kovacec, Ana Cristina M. Rosa, Dina Maria L. F. S. Loff, Joaquim J. A. Júdice, José Augusto M. Ferreira, Maria C. A. Gouveia e Maria de Nazaré S. Q. M. Lopes, pelas palavras de elogio e incentivo frequentemente proferidas.*

*Às doutoras Elisa Motta, Carolina Pinto e Maria Lídia Cardoso, por, no exercício das suas funções, não pouparem esforços para que o meu sonho e de muitos outros estudantes tornassem realidades.*

*Aos meus pais, Cipriana e Miguel, a quem também dedico este trabalho, por iluminarem em mim o caminho da vida. A estes, não me parece que haja uma lista de adjectivos que aqui possa apresentar que caracterize a minha profunda gratidão.*

*Aos meus irmãos e aos meus primos, pelo amor perpétuo que nunca hesitaram em demonstrar, encorajando-me sempre nas minhas ousadas aventuras, estas que nunca cessarão.*

*Aos meus tios, por serem o meu suporte na ausência dos meus pais, em especial, Manuel Silva que em boas graças acolheu-me no seu lar e que, desde a minha chegada a Portugal, tem-me apoiado incondicionalmente e confiado um amor que causa inveja a qualquer filho.*

*A todos os meus amigos, em especial, Leo, Lili, Meno, Ny e Ro, pela camaradagem e cumplicidade singulares reveladas durante toda esta caminhada neste terreno íngreme e escorregadio, que é a nossa vida académica.*



# Conteúdo

<b>1</b>	<b>Introdução e algumas noções</b>	<b>1</b>
1.1	A abordagem paramétrica . . . . .	2
1.2	A abordagem não paramétrica . . . . .	3
1.3	Algumas definições . . . . .	5
<b>2</b>	<b>Estimadores não paramétricos habituais</b>	<b>7</b>
2.1	O regressograma . . . . .	7
2.2	O Nadaraya-Watson . . . . .	12
<b>3</b>	<b>Estimação com observações em falta</b>	<b>23</b>
3.1	Introdução . . . . .	23
3.2	Estimador de Cheng para a média - Modelo não paramétrico . . . . .	25
3.3	Estimador para a média - Modelo parcialmente linear . . . . .	32
3.3.1	Estimação da função de regressão . . . . .	33
3.3.2	Uma classe geral de estimadores . . . . .	35
3.4	Eficiência dos estimadores . . . . .	44
3.5	Nota conclusiva . . . . .	47



# Capítulo 1

## Introdução e algumas noções

Os modelos de regressão são dos mais populares nos diversos domínios de aplicação da estatística. De facto, eles permitem introduzir variações aleatórias num modelo que descreve uma ou várias variáveis à custa de algum outro conjunto de variáveis explicativas. É possível assim ir para além das relações funcionais analíticas, incorporando não só efeitos aleatórios mas, principalmente, algum desconhecimento das relações existentes entre as diversas variáveis que intervêm no modelo. Para introduzir algum formalismo que permite algum tratamento matemático, representemos por  $Y$  a variável aleatória que pretendemos explicar e por  $X$  um vector de  $d$  variáveis aleatórias que são explicativas, isto é, que de alguma forma determinam o valor de  $Y$ . Do ponto de vista matemático  $Y$  poderia também ser um vector, mas iremos ocupar-nos neste texto apenas do caso unidimensional. Para o estudo subsequente neste trabalho iremos assumir que a relação entre estas variáveis é da seguinte forma:

$$Y = m(X) + \varepsilon, \tag{1.1}$$

onde  $\varepsilon$  é uma variável aleatória de média nula e variância finita, que designaremos por *erro*. Desta forma,  $Y$  torna-se uma variável aleatória cuja média é  $m(X)$  e cuja variância é  $Var(\varepsilon)$ . A função  $m$ , que é habitualmente desconhecida, será designada por *função de regressão*.

Para descrever convenientemente as propriedades do modelo de regressão (1.1) é necessário recorrer a alguns conceitos probabilísticos. Em face de dados reais provenientes do modelo de regressão, conhecem-se apenas alguns pares de pontos  $(y_1, x_1), \dots, (y_n, x_n)$ , onde  $(y_i, x_i) \in (\mathbb{R}, \mathbb{R}^d)$ ,  $i = 1, \dots, n$ , e procura-se uma aproximação para a *função de regressão*  $m$  que explica a relação entre as variáveis. Atendendo a que a relação inclui uma componente aleatória é conveniente interpretar estes dados como uma concretização de um vector aleatório. Mais ainda, para a descrição dos aspectos formais do modelo que iremos estudar, podemos admitir que os pontos  $x_1, \dots, x_n$  onde são feitas as observações são também eles aleatórios. Isto é, admite-se que os pontos de observação que deram origem aos dados não são bem determinados

à partida. Este ponto de vista altera ligeiramente o quadro formal de definição do modelo. A relação (1.1) mantém-se, mas  $m$  passa agora a ser interpretada como o valor esperado quando é dado que  $X$  assumiu um valor particular. Isto é, matematicamente a *função de regressão* passa a representar a esperança condicional de  $Y$  quando é conhecido o valor da variável ou vector explicativo:  $m(x) = E(Y|X = x)$  ou ainda, de forma mais curta,  $m(X) = E(Y|X)$ .

Dada a natureza aleatória do modelo descrito, abordemos a construção de aproximações para a *função de regressão*  $m$  do ponto de vista estatístico. Assim, seja  $\{(Y_i, X_i)\}_{i=1}^n$ ,  $n \in \mathbb{N}$ , uma sucessão de observações, onde cada par  $(Y_i, X_i)$  é independente e identicamente distribuído (*i.i.d.*) com  $(Y, X)$ , isto é, uma amostra para o par  $(Y, X)$ . É a partir desta amostra que pretendemos construir e estudar as aproximações para a função  $m$ .

Nas secções seguintes far-se-á uma breve apresentação de metodologias de abordagem, tendo em conta a inclusão no modelo de hipóteses sobre a função  $m$ . Em capítulos posteriores, centrar-nos-emos na abordagem não paramétrica, pois é desta que pretendemos apresentar neste trabalho uma extensão ao caso em que a amostra não é totalmente conhecida.

### 1.1. A abordagem paramétrica

Uma primeira abordagem simplificadora ao problema da aproximação da *função de regressão* consiste em admitir que conhecemos a forma da função  $m$ , isto é, que a menos da explicitação de alguns parâmetros identificadores, conhecemos a família de funções à qual  $m$  pertence. Formalmente, seja  $\Theta$  uma família de funções em que cada elemento é identificado à custa de um parâmetro  $\beta$  (possivelmente multidimensional):

$$\Theta = \{m(\cdot, \beta), \beta \in \Upsilon\},$$

onde  $\Upsilon \subset \mathbb{R}^a$ . Neste caso, o modelo (1.1) reescreve-se na forma

$$Y = m(X, \beta) + \varepsilon, \tag{1.2}$$

e há que procurar o valor de  $\beta$  que melhor explica os dados observados. Para levar a cabo esta procura é necessário introduzir um critério de erro que depois será optimizado para obter a melhor aproximação possível para o parâmetro. Um dos critérios de erro mais utilizados é o que de seguida definimos.

**Definição 1.1.** Chama-se *erro quadrático médio (EQM)* associado à escolha do parâmetro  $\beta$  a

$$T(\beta) = \sum_{i=1}^n (Y_i - m(X_i, \beta))^2.$$

**Nota 1.1.** Em rigor a função  $T$  agora introduzida depende também da amostra. Como iremos proceder à construção de aproximações perante uma dada amostra, não escrevemos explicitamente esta dependência para não tornar a notação demasiado pesada.

Uma definição mais alargada de *EQM* é introduzida um pouco mais à frente (Definição 1.3).

**Definição 1.2.** Define-se o estimador para o parâmetro  $\beta$  como o minimizante de  $T(\beta)$ , isto é,  $\hat{\beta} = \operatorname{argmin}_{\beta} T(\beta)$ .

**Exemplo 1.1.** Se o modelo de regressão (1.2) assume a forma  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$ , então o parâmetro minimizante do *EQM* será a solução do problema de minimização

$$\min_{\beta} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2), \quad (1.3)$$

onde  $\beta = (\beta_0, \beta_1, \beta_2)$ .

Ao modelo de regressão (1.1) assumindo que a *função de regressão* admite a representação (1.2) chamaremos *modelo de regressão paramétrico (MRP)*. Com esta simplificação sobre a representação da *função de regressão*, reduzimos o problema da construção dum aproximação à optimização de uma função com  $a$  (a dimensão do parâmetro que caracteriza  $\beta$ ) variáveis.

## 1.2. A abordagem não paramétrica

Uma das dificuldades intrínsecas ao método descrito na secção anterior é o risco de uma má escolha da família  $\Theta$  a que, supostamente, pertence a *função de regressão*. O problema matemático de optimização terá, naturalmente, uma solução, mas esta está destinada a fornecer más aproximações no que diz respeito à descrição de observações. Por outro lado, a disposição dos dados observados pode revelar uma estrutura muito complexa não sugerindo nenhuma família de funções conhecidas. Nestes casos, é conveniente retirar a restrição sobre a forma funcional de  $m$  e “deixar os dados mostrarem qual a curva que melhor se lhe ajusta”, ou seja, assume-se que  $m$  é uma

função completamente desconhecida e usam-se os dados para inferir sobre a sua forma. Nestas condições diz-se que estamos a estudar um *modelo de regressão não paramétrico* (*MRNP*). Este modelo suscitou o desenvolvimento de vários estimadores (não paramétricos) para  $m$ , entre os quais os mais populares são o *regressograma* (consultar Bosq et Lecoutre [?]), pela sua simplicidade, e o *Nadaraya-Watson* (*N-W*) que serão descritos no próximo capítulo. Mais detalhes acerca do estimador *N-W* podem ser encontrados em Härdle [?], Wand and Jones [?] e Härdle, Müller, Sperlich and Werwatz [?], entre outras referências.

A construção do estimador pode ser abordada através da optimização de uma função do tipo (1.2). No entanto, não existe nenhum parâmetro relativamente ao qual optimizar, pelo que a minimização terá que ser feita relativamente à escolha de função  $m$ , pelo que se trata de uma optimização em dimensão infinita.

Estes modelos não paramétricos contêm uma dificuldade devida à dimensão do espaço em que toma valores o vector aleatório  $X$ . Intuitivamente, se a dimensão deste espaço for grande, a quantidade de pontos necessária para obter uma caracterização razoável das aproximações deverá ser também bastante grande. A ideia subjacente é que num espaço de dimensão grande, o espaço vazio abunda e, onde não há informação disponível também não é possível construir aproximações com alguma qualidade. A este problema refere-se usualmente por *maldição da dimensionalidade*.

Podemos ainda construir modelos em que os *MRP* e *MRNP* se complementem, na medida em que, como no modelo a estudar no Capítulo 3, sendo a dimensão do vector aleatório  $X$  relativamente grande, há conveniência em combinar técnicas das duas abordagens para reduzir a dimensão do espaço em que se aplica a abordagem não paramétrica, ganhando com isso um melhor desempenho dos estimadores. Pode ainda acontecer que parte de  $X$  revele uma estrutura muito simples, podendo ser modelada parametricamente e outra parte demasiada complexa, pelo que necessita de uma modelação não paramétrica. Às abordagens que envolvem combinações de técnicas paramétricas e não paramétricas chamaremos de *regressão semiparamétrica*. Neste caso um modelo comum é  $m(X) = f(X^*, \beta) + m^*(T)$ , onde  $f(X^*, \beta)$  é a parte paramétrica e  $m^*(T)$  a parte não paramétrica, com  $X = (X^*, T)$ . Para o estudo que pretendemos realizar neste trabalho, introduziremos, no último capítulo, os *modelos de regressão parcialmente linear* (*MRPL*), constituindo estes um caso particular dos *modelos de regressão semiparamétrica*, em que a parte paramétrica se supõe ser linear. Para um estudo mais geral sobre os *MRPL* consultar Härdle, Liang



and Gao [?].

### 1.3. Algumas definições

Nesta secção introduzimos algumas noções genéricas úteis para o desenvolvimento deste trabalho.

**Definição 1.3.** Seja  $\hat{\theta}_n = \gamma(X_1, \dots, X_n)$  um estimador de  $\theta = \gamma^*(X)$ , onde  $\theta$  é um parâmetro desconhecido ( $E(X)$ ,  $Var(X)$ , ...). Diz-se que  $\hat{\theta}_n$  converge em média quadrática (m.q.) para  $\theta$  se

$$E\left(\hat{\theta}_n - \theta\right)^2 \longrightarrow 0$$

quando  $n \rightarrow +\infty$ . A  $E\left(\hat{\theta}_n - \theta\right)^2$  chama-se *EQM* associado ao estimador  $\hat{\theta}_n$  e denota-se por  $EQM(\hat{\theta}_n)$ .

Notemos que o *EQM* associado a  $\hat{\theta}_n$  pode ainda ser reescrito da seguinte forma:

$$\begin{aligned} EQM\left(\hat{\theta}_n\right) &= E\left(\hat{\theta}_n - \theta\right)^2 \\ &= Var\left(\hat{\theta}_n\right) + \left(E\left(\hat{\theta}_n\right) - \theta\right)^2. \end{aligned} \quad (1.4)$$

**Definição 1.4.** Diz-se que  $\hat{\theta}_n$  converge em probabilidade para  $\theta$  e denota-se por  $\hat{\theta}_n \xrightarrow{P} \theta$  se,

$$\forall \eta > 0, \lim_{n \rightarrow +\infty} P(|\hat{\theta}_n - \theta| > \eta) = 0.$$

**Definição 1.5.**

- Diz-se que  $\hat{\theta}_n$  é um estimador (de  $\theta$ ) consistente em m.q. se, quando  $n \rightarrow +\infty$ ,

$$EQM\left(\hat{\theta}_n\right) \longrightarrow 0;$$

- Diz-se que  $\hat{\theta}_n$  é um estimador (de  $\theta$ ) consistente em probabilidade se, quando  $n \rightarrow +\infty$ ,

$$\hat{\theta}_n \xrightarrow{P} \theta.$$

**Nota 1.2.** Uma vez que a convergência em m.q. implica convergência em probabilidade (teorema 5.12 de Gonçalves e M.Lopes [?]), a consistência em m.q. implica consistência em probabilidade.

**Definição 1.6.** Seja  $\xi$  uma função de  $n \in \mathbb{N}$ , tal que  $\xi(n) > 0$ .

- Diz-se que uma quantidade  $a_n$  é um  $o(\xi(n))$  se  $\lim_{n \rightarrow +\infty} \frac{a_n}{\xi(n)} = 0$ ;
- Diz-se que uma quantidade  $a_n$  é um  $o_p(\xi(n))$  se  $\frac{a_n}{\xi(n)} \xrightarrow{P} 0$ .



# Capítulo 2

## Estimadores não paramétricos habituais

De entre os vários estimadores consistentes para a *função de regressão* desenvolvidos na sequência do estudo do *MRNP*, neste trabalho, estamos particularmente interessados em introduzir o *regressograma* e o *N-W*, não so pelos seus óptimos desempenhos, conforme poderemos constatar nas duas secções seguintes, mas também por exercerem um papel de extrema importância no estudo que pretendemos desenvolver mais a diante, sobretudo o *N-W*.

Nas secções seguintes, salvo menção contrária, vamos assumir que  $X$  é unidimensional, deixando o caso genérico,  $X \in \mathbb{R}^d$ , para o Capítulo 3.

Para, num primeiro contacto, transmitir uma ideia mais sugestiva sobre os papéis dos estimadores não paramétricos para a *função de regressão*, comecemos por introduzir o *regressograma*.

### 2.1. O regressograma

Para esta secção, assumamos, sem perda de generalidade, que a variável aleatória real (*v.a.r.*)  $X$  toma valores no intervalo  $[0, 1]$ . Pois, existem várias aplicações que asseguram a bijectividade entre o conjunto  $\mathbb{R}$  e o intervalo  $]0, 1[$ , nomeadamente  $\frac{1}{\pi} \arccotan(\cdot)$ . A inclusão dos pontos 0 e 1 deve-se exclusivamente ao facto do intervalo  $[0, 1]$  ser compacto, tendo portanto uma estrutura mais adequada para transmitir a ideia subjacente ao *regressograma*.

Seja  $x$  uma concretização da *v.a.r.*  $X$ . O *regressograma* usa as observações de  $X$  que estão numa certa vizinhança contendo  $x$  para inferir sobre a curva desconhecida,  $m$ , nessa mesma vizinhança. Mais concretamente e de um modo geral, este estimador atribui à curva desconhecida uma mesma estimativa em todos os pontos numa mesma vizinhança (previamente fixada), pelo que desde já podemos dizer que este é muito restritivo. Pois, desta forma, a curva desconhecida será “forçada” a assumir uma mesma estimativa em todos os pontos de cada vizinhança.

Após esta breve exposição, resta-nos portanto formalizar a noção de *regressograma*. Neste sentido, consideremos uma sucessão de números naturais  $k = k(n) \geq 2$ , onde  $k \rightarrow +\infty$  quando  $n \rightarrow +\infty$ , e  $k$  sub-intervalos  $\pi_{kq} = \left[\frac{q-1}{k}, \frac{q}{k}\right]$ ,  $q = 1, \dots, k$ , do intervalo  $[0, 1]$  (de amplitudes iguais a  $h = \frac{1}{k}$ ), pelo que  $[0, 1] = \cup_{q=1}^k \pi_{kq}$ . Consideremos ainda  $I_{nq} = \{i : 1 \leq i \leq n, X_i \in \pi_{kq}\}$ , o conjunto dos índices  $i$ 's para o quais  $X_i$ 's estão em  $\pi_{kq}$ , e  $\eta_{nq} = |I_{nq}|$ , o cardinal do conjunto  $I_{nq}$ .

**Definição 2.1.** O *regressograma* associado à sucessão de observações  $\{(Y_i, X_i)\}_{i=1}^n$ , sobre  $\pi_{kq}$ , é o estimador  $\widehat{m}_r$ , onde

$$\widehat{m}_r(x) = \begin{cases} \frac{1}{\eta_{nq}} \sum_{i \in I_{nq}} Y_i & \text{se } \eta_{nq} \neq 0 \\ 0 & \text{se } \eta_{nq} = 0 \end{cases}, \quad \forall x \in \pi_{kq}. \quad (2.1)$$

Estamos a adoptar a convenção  $\frac{s}{0} = 0$ , para qualquer ponto  $s$  no conjunto  $\mathbb{R}$ .

Notemos que, embora assintoticamente eficiente conforme a proposição que se segue, este estimador atribui a  $m$  o mesmo valor para todos os pontos do sub-intervalo  $\pi_{kq}$  ( $q = 1, \dots, k$ ), originando assim uma curva estimada em patamares e, conseqüentemente, descontínua, o que vem confirmar a sua restritividade.

Para enunciar convenientemente um resultado que assegura a consistência deste estimador, o *regressograma*, definamos  $v(x) = Var(Y|X = x)$  a variância condicional de  $Y$  dado  $X = x$  e consideremos as seguintes hipóteses:

(H<sub>1</sub>)  $E(Y^2) < \infty$ ;

(H<sub>2</sub>)  $v$  e  $m$  são funções contínuas em  $x \in \pi_{kq}$ ;

(H<sub>3</sub>)  $X$  admite função densidade de probabilidade, diga-se  $f$ , contínua, não nula em  $x \in \pi_{kq}$ ;

(H<sub>4</sub>)  $h \rightarrow 0$  e  $nh \rightarrow +\infty$ .

**Proposição 2.1.** *Assumindo as hipóteses H<sub>1</sub> a H<sub>4</sub> acima listadas, tem-se*

$$EQM(\widehat{m}_r(x)) \rightarrow 0, \quad \forall x \in \pi_{kq}.$$

*Demonstração.* A equação (1.4) permite-nos escrever

$$EQM(\widehat{m}_r(x)) = Var(\widehat{m}_r(x)) + (E(\widehat{m}_r(x)) - m(x))^2. \quad (2.2)$$

Comecemos por provar que o segundo termo do lado direito da equação anterior,  $(E(\widehat{m}_r(x)) - m(x))^2$ , converge para 0. Com efeito,

$$\begin{aligned} E(\widehat{m}_r(x)) &= \sum_{m=0}^n E(\widehat{m}_r(x) | \eta_{nq} = m) P(\eta_{nq} = m) \\ &= E(Y | X \in \pi_{kq}) \sum_{m=1}^n P(\eta_{nq} = m) \\ &= E(Y | X \in \pi_{kq}) [1 - (1 - P(X \in \pi_{kq}))^n]. \end{aligned}$$

Atendendo a que  $f$  é uma função densidade de probabilidade contínua e não nula em  $x$ , à continuidade de  $m$  em  $x$  e como  $nh \rightarrow +\infty$ , tem-se o seguinte:

$$\begin{aligned} P(X \in \pi_{kq}) &= \int_{\pi_{kq}} (f(x) + f(t) - f(x)) dt \\ &= f(x)h + \int_{\pi_{kq}} (f(t) - f(x)) dt \\ &= hf(x)(1 + o(1)), \end{aligned} \tag{2.3}$$

$$\begin{aligned} E(Y | X \in \pi_{kq}) &= \frac{1}{P(X \in \pi_{kq})} \int_{\pi_{kq}} (m(x) + m(t) - m(x)) f(t) dt \\ &= m(x) + \frac{1}{P(X \in \pi_{kq})} \int_{\pi_{kq}} (m(t) - m(x)) f(t) dt \\ &= m(x) (1 + o(1)) \end{aligned} \tag{2.4}$$

e

$$\begin{aligned} (1 - P(X \in \pi_{kq}))^n &= \left(1 - \frac{nhf(x)(1 + o(1))}{n}\right)^n \\ &= e^{-nhf(x)(1+o(1))} + o(1). \end{aligned} \tag{2.5}$$

Assim, por (2.3), (2.4) e (2.5), tem-se

$$E(\widehat{m}_r(x)) = \left(m(x) - m(x)e^{-nhf(x)}\right) (1 + o(1)) \tag{2.6}$$

e, portanto,

$$(E(\widehat{m}_r(x)) - m(x))^2 \rightarrow 0. \tag{2.7}$$

De seguida, tendo em conta a igualdade

$$Var(\widehat{m}_r(x)) = E(\widehat{m}_r(x))^2 - E^2(\widehat{m}_r(x)),$$

provaremos que  $Var(\widehat{m}_r(x))$  converge para 0. Ora,

$$\begin{aligned}
 E(\widehat{m}_r(x))^2 &= \sum_{m=0}^n E((\widehat{m}_r(x))^2 | \eta_{mq} = m) P(\eta_{mq} = m) \\
 &= \sum_{m=1}^n \{Var(\widehat{m}_r(x) | \eta_{mq}) + E^2(\widehat{m}_r(x) | \eta_{mq} = m)\} P(\eta_{mq} = m) \\
 &= Var(Y | X \in \pi_{kq}) \sum_{m=1}^n \frac{1}{m} P(\eta_{mq} = m) \\
 &\quad + E^2(Y | X \in \pi_{kq}) \sum_{m=1}^n P(\eta_{mq} = m) \\
 &= Var(Y | X \in \pi_{kq}) S_n + B_n,
 \end{aligned} \tag{2.8}$$

onde  $S_n = \sum_{m=1}^n \frac{1}{m} P(\eta_{mq} = m)$  e, por (2.3) e (2.4),

$$\begin{aligned}
 B_n &= E^2(Y | X \in \pi_{kq}) \sum_{m=1}^n P(\eta_{mq} = m) \\
 &= E^2(Y | X \in \pi_{kq}) (1 - P(\eta_{mq} = 0)) \\
 &= E^2(Y | X \in \pi_{kq}) (1 - [1 - P(X \in \pi_{kq})]^n) \\
 &= m^2(x) \left(1 - e^{-nhf(x)}\right) (1 + o(1)).
 \end{aligned} \tag{2.9}$$

Para  $S_n$ , notemos que  $\eta_{mq}$  segue a lei binomial de parâmetros  $n$  e  $p = P(X \in \pi_{kq}) = hf(x)(1 + o(1))$ . Deste modo,

$$\begin{aligned}
 S_n &= \sum_{m=1}^n \frac{1}{m} \binom{n}{m} p^m (1-p)^{n-m} \\
 &= \frac{1}{(n+1)p} \sum_{m=1}^n \frac{m+1}{m} \binom{n+1}{m+1} p^{m+1} (1-p)^{n-m} \\
 &= \frac{1}{np} (1 + o(1)) + \frac{1}{(n+1)p} \sum_{m=1}^n \frac{1}{m} \binom{n+1}{m+1} p^{m+1} (1-p)^{n-m} \\
 &= \frac{1}{nhf(x)} (1 + o(1)) + b_n.
 \end{aligned} \tag{2.10}$$

O termo  $b_n$  é majorado por um termo que, nas condições das hipóteses, converge para 0. De facto,

$$\begin{aligned}
 b_n &= \frac{1}{(n+1)p} \sum_{m=1}^n \frac{1}{m} \binom{n+1}{m+1} p^{m+1} (1-p)^{n-m} \\
 &= \frac{1}{(n+1)(n+2)p^2} \sum_{m=1}^n \frac{m+2}{m} \binom{n+2}{m+2} p^{m+2} (1-p)^{n-m} \\
 &\leq \frac{3}{(n+1)(n+2)p^2} \sum_{m=1}^n \binom{n+2}{m+2} p^{m+2} (1-p)^{n-m} \\
 &= \frac{3}{(n+1)(n+2)p^2} (1 + o(1)) \\
 &= o(1).
 \end{aligned} \tag{2.11}$$

Finalmente, por (2.6), (2.8), (2.9), (2.10) e (2.11), concluímos que  $Var(\widehat{m}_r(x)) \rightarrow 0$ . Assim, provamos que  $EQM(\widehat{m}_r(x)) \rightarrow 0$ , conforme se pretendia.  $\square$

A demonstração anterior pode ser encontrada em Bosq et Lecoutre [?].

**Exemplo 2.1** (Aplicação do regressograma).

Começemos por supor que a relação entre as *v.a.r.*  $X$  e  $Y$  é descrita pela equação  $Y = m(X) + \varepsilon$ , onde  $m(X) = \sqrt[3]{X}$  e  $\varepsilon$  é o erro. Desta forma, a curva “desconhecida” que pretendemos estimar é  $m(X) = \sqrt[3]{X}$ , com  $X$  no intervalo  $[0, 1]$ . Consideremos agora o modelo de regressão

$$Y = \sqrt[3]{X} + \varepsilon, \quad (2.12)$$

supondo que o erro,  $\varepsilon$ , é uma *v.a.r.* seguindo a *lei normal* de média nula e variância 0.1, e usemos o regressograma para construir uma curva aproximada para  $m$ . Para isto, consideramos, por exemplo,  $k = 30$ , pelo que  $h = \frac{1}{30}$ , geramos 100 observações da variável  $X$ , uniformemente distribuídas no intervalo  $[0, 1]$ , e 100 observações do erro, pelo que obtivemos uma sucessão aleatória de observações  $\{Y_i\}_{i=1}^{100}$ , onde  $Y_i = \sqrt[3]{X_i} + \varepsilon_i$ . Tais observações encontram-se dispostas no gráfico da Figura 2.1.

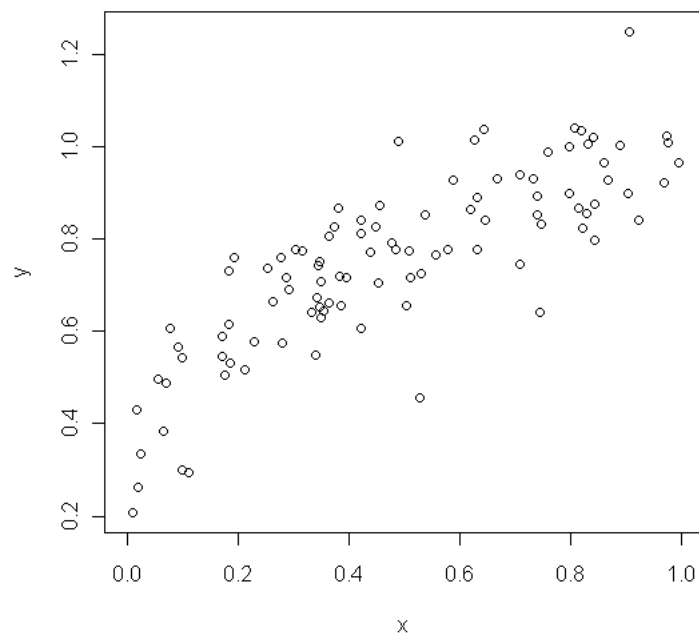


Figura 2.1: Disposição (gráfica) de 100 observações do modelo (2.12).

A representação gráfica da curva (que pretendemos aproximar)  $m$  e da curva aproximada (utilizando o estimador regressograma), vem realçar a restritividade do

*regressograma* (ver a Figura 2.2).

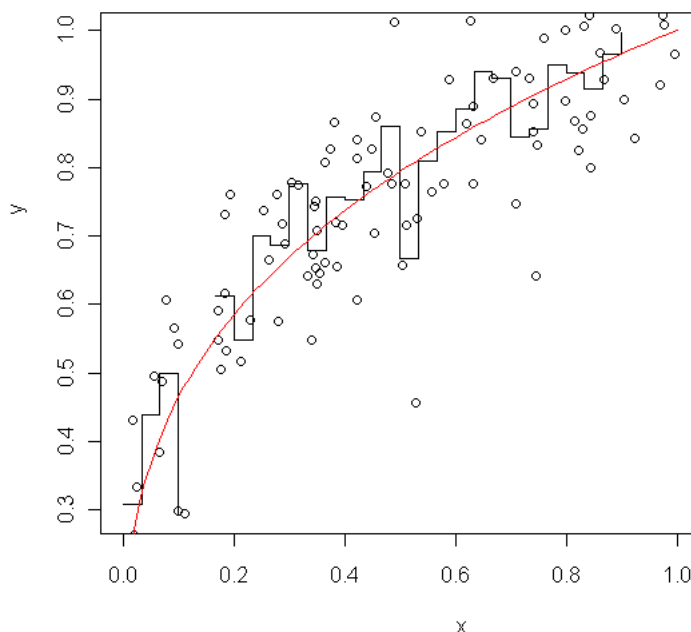


Figura 2.2: *Representação gráfica da curva  $m$  e do regressograma. A curva  $m$  encontra-se a **vermelho** e o regressograma a **preto**.*

Com efeito, mesmo sem nenhum conhecimento a priori sobre  $m$ , espera-se que esta seja uma curva um tanto quanto suave (regular), o que obviamente, não é salvaguardado pelo *regressograma* que, conforme se pode constatar, imputa-lhe uma curva estimada em patamares e, logo, descontínua. Contudo, apesar das suas características, o *regressograma* permite-nos ter uma noção da forma da verdadeira curva, sem nenhuma assunção a priori sobre esta mesma curva.

## 2.2. O Nadaraya-Watson

Conforme podemos observar na secção anterior, o *regressograma* é um estimador relativamente fácil de construir, mas, conforme veremos no decorrer desta secção, isto não é o caso do estimador *N-W*.

Começemos a construção do *N-W* partindo do conhecimento do tipo de estimador para densidade que usaremos, o *estimador do núcleo (kernel) para densidade*.

Ora, suponhamos então que as variáveis aleatórias  $Y$  e  $X$  admitem função densidade de probabilidade (*fdp*) conjunta  $g$  (desconhecida) e consideremos a expressão da *função de regressão*,  $m(X) = E(Y|X)$ .



Observando a expressão de  $m$  em  $X = x$ ,

$$m(x) = E(Y|X = x) = \int yg^x(y)dy = \frac{1}{f(x)} \int yg(y, x)dy, \quad (2.13)$$

onde  $g^x$  é a *fdp* condicional de  $Y$  dado  $X = x$  e  $f$  *fdp* marginal de  $X$ , concluímos que, dado a sucessão de observações  $\{(Y_i, X_i)\}_{i=1}^n$  do par  $(Y, X)$ ,  $m(x)$  depende apenas das quantidades (desconhecidas)  $f(x)$  e  $g(y, x)$ .

Estimemos as quantidades  $f(x)$  e  $g(y, x)$ , usando o *estimador de kernel para densidade*. Para um estudo mais detalhado sobre a estimação de *fdp*, consultar Härdle, Müller, Sperlich and Werwatz [?] ou Wand and Jones [?].

Portanto, uma estimativa para  $f(x)$  será

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right), \quad (2.14)$$

onde  $K$  é uma *fdp* simétrica sobre  $\mathbb{R}$  tal que  $\int sK(s)ds = 0$ , usualmente denominada *núcleo (kernel)*, e  $h = h(n)$  é uma sucessão a decrescer para 0 enquanto  $n \rightarrow +\infty$ . Para uma amostra da variável  $X$  de tamanho  $n$ , a função *kernel* atribui a cada observação  $X_i$ ,  $i = 1, \dots, n$ , numa vizinhança  $h$  de  $x$ , um peso inversamente proporcional a  $h$ ,  $\frac{1}{nh}K\left(\frac{x-X_i}{h}\right)$ , na estimação de  $f(x)$ , pelo que quanto mais próximo de  $x$  esta se encontrar maior será o seu peso. Na Tabela 2.1 estão algumas das funções *kernel* mais habituais.

Prova-se que o estimador de *kernel* para  $f(x)$ ,  $\widehat{f}_n(x)$ , é *consistente em m.q.* e que, das funções *kernel* conhecidas, o Epanechnikov é a que minimiza o *EQM*  $\left(\widehat{f}_n(x)\right)$ . Tais provas podem ser encontradas em Härdle, Müller, Sperlich and Werwatz [?] ou Wand and Jones [?].

Kernel	K(s)
Uniforme	$\frac{1}{2}I_{\{ s \leq 1\}}$
Triangular	$(1 -  s )I_{\{ s \leq 1\}}$
Epanechnikov	$\frac{3}{4}(1 - s^2)I_{\{ s \leq 1\}}$
Quártico	$\frac{15}{16}(1 - s^2)^2I_{\{ s \leq 1\}}$
Tricúbico	$\frac{35}{32}(1 - s^2)^3I_{\{ s \leq 1\}}$
Gaussiano	$\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}$

Tabela 2.1: Algumas funções *kernel*.  $I_{\{\cdot\}}$  é a função indicatriz

Para  $g(y, x)$ , uma estimativa será

$$\hat{g}_n(y, x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{hl} \mathcal{K} \left( \frac{y - Y_i}{l}, \frac{x - X_i}{h} \right), \quad (2.15)$$

onde  $l = l(n)$ , à semelhança de  $h$ , é uma sucessão a decrescer para 0 enquanto  $n \rightarrow +\infty$ ,  $\mathcal{K}$  denota a função *kernel* operando sobre argumento de dimensão  $d$ , usualmente denominada *kernel multidimensional*.

Vejamos duas formas simples de construir um *kernel multidimensional* a partir da função *kernel* operando sobre argumento unidimensional,  $K$ , este que designaremos por *kernel unidimensional*:

(a<sub>1</sub>) *kernel produtório (multiplicativo)*  $\mathcal{K}^p(s_1, \dots, s_d) = \prod_{i=1}^d K(s_i)$  que, conforme se pode constatar, é simplesmente o produto dos  $d$  valores do *kernel unidimensional* operando sobre cada um dos  $d$  argumentos unidimensionais  $s_i$ ,  $i = 1, \dots, d$ ;

(a<sub>2</sub>) *kernel esférico*  $\mathcal{K}^e(s) = \frac{1}{\int K(\|s\|)} K(\|s\|)$ , onde  $\|s\|$  é a norma euclidiana de  $s = (s_1, \dots, s_d)$ . Esta função deve o seu nome ao facto de tomar o mesmo valor em todos os pontos  $s$  numa esfera centrada em 0 e de raio 1. Obviamente, no caso de  $s$  ser bidimensional, deveremos falar em círculo em vez de esfera.

Para facilitar os cálculos é mais conveniente assumir  $\mathcal{K}(s) = \mathcal{K}^p(s)$ , sendo que neste caso o estimador de *kernel* para  $g$  no ponto  $(y, x)$ ,  $\hat{g}_n(y, x)$ , assume a forma

$$\hat{g}_n(y, x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{hl} K \left( \frac{y - Y_i}{l} \right) K \left( \frac{x - X_i}{h} \right). \quad (2.16)$$

Consideremos a função  $K_h$  definida da seguinte forma:

$$K_h(u) = \frac{1}{h} K \left( \frac{u}{h} \right).$$

Definida desta forma,  $K_h$  é usualmente designada por função *kernel rescalada*. Pode-se constatar facilmente que, sendo  $K$  uma *fdp* simétrica sobre  $\mathbb{R}$  tal que  $\int sK(s)ds = 0$ ,  $K_h$  também será.

**Definição 2.2.** Define-se a função  $W_{hi}$ , usualmente designada por *peso*, associado à sucessão de observações  $\{X_i\}_{i=1}^n$  de  $X$  por

$$W_{hi}(s) = \frac{K_h(s - X_i)}{\frac{1}{n} \sum_{j=1}^n K_h(s - X_j)}, \quad i = 1, \dots, n. \quad (2.17)$$

**Definição 2.3.** O estimador  $N$ - $W$  associado à sucessão de observações  $\{(Y_i, X_i)\}_{i=1}^n$ , do par  $(Y, X)$ , é o estimador  $\widehat{m}_{NW}$  de  $m$  em  $x$ , onde

$$\begin{aligned}\widehat{m}_{NW}(x) &= \frac{1}{\widehat{f}_n(x)} \int y \widehat{g}_n(y, x) dy \\ &= \frac{1}{n} \sum_{i=1}^n \left( \frac{K_h(x - X_i)}{\frac{1}{n} \sum_{j=1}^n K_h(x - X_j)} \right) Y_i \\ &= \frac{1}{n} \sum_{i=1}^n W_{hi}(x) Y_i\end{aligned}\tag{2.18}$$

se  $\widehat{f}_n(x) \neq 0$  e  $\widehat{m}_{NW}(x) = 0$  se  $\widehat{f}_n(x) = 0$  e  $W_{hi}$  a função *peso*.

Notemos que se o domínio da função  $K$  for limitado, o que, exceptuando o *kernel* gaussiano, é o caso de todas as funções *Kernel* na Tabela 2.1,  $\widehat{f}_n(x) = 0$  se verifica quando nenhuma das observações de  $X$  estiver numa vizinhança  $h$  de  $x$ , o que acontece quando as observações forem muito esparsas. Notemos ainda que, quando  $X$  não é unidimensional, o peso  $W_{hi}(x)$  na expressão do estimador  $N$ - $W$  (equação (2.18)) assume a forma

$$W_{hi}(x) = \frac{\mathcal{K}\left(\frac{x_1 - X_{i1}}{h_1}, \dots, \frac{x_d - X_{id}}{h_d}\right)}{\frac{1}{n} \sum_{j=1}^n \mathcal{K}\left(\frac{x_1 - X_{j1}}{h_1}, \dots, \frac{x_d - X_{jd}}{h_d}\right)},\tag{2.19}$$

onde  $h_i$ 's ( $i = 1, \dots, n$ ) são sucessões a decrescerem para 0 enquanto  $n \rightarrow +\infty$ . Neste caso, o desempenho do estimador pode ser negativamente afectado pela *maldição da dimensionalidade*. Pois, intuitivamente, relativamente à construção de aproximações, quanto maior for a dimensão do espaço onde toma valores a variável em estudo ( $X$ , neste caso), mais insignificantes tornam-se as informações representadas pelos pontos observados e, conseqüentemente, menor será a fiabilidade das aproximações construídas à custa destes pontos.

Supondo que  $\widehat{m}_{NW}$  está definido somente nos pontos  $x = X_i$ ,  $i = 1, \dots, n$ , com  $n$  fixo, tem-se que  $\widehat{m}_{NW}(x) \rightarrow Y_i$ , à medida que  $h$  for se aproximando de 0, isto é, uma interpolação das observações. Por outro lado, escolhendo  $h$  suficientemente grande, tem-se  $\widehat{m}_{NW}(x) = \frac{1}{n} \sum_{i=1}^n Y_i$ .

Para adaptar o  $N$ - $W$  ao caso em que a variável  $X$  não é aleatória, basta usar o facto da sua *fdp*,  $f$ , ser conhecida. Deste modo, não sendo necessário estimar  $f(x)$ , um estimador consistente para a *função de regressão* será  $\widehat{m}_{NW}^*$ , onde

$$\widehat{m}_{NW}^*(x) = \frac{1}{n} \sum_{i=1}^n W_{hi}^*(x) Y_i$$

se  $f(x) \neq 0$  e  $\widehat{m}_{NW}^*(x) = 0$  se  $f(x) = 0$ , com  $W_{hi}^*(x) = \frac{K_h(x-X_i)}{f(x)}$ . Neste caso,  $m(x)$  depende apenas da quantidade  $g(y, x)$  e portanto mais fácil de estimar.

A título de exemplo, consideremos uma experiência onde é suposto explicar o *tempo de reação* ( $Y$ ) a um incidente no tráfico rodoviário de um indivíduo, em função do *consumo pessoal de cerveja* ( $X$ ). De facto, aqui, antes de conduzir a experiência, escolhe-se como observar  $X$ , ou seja, que quantidade ( $x_i, i = 1, \dots, n$ ) de cerveja atribuir a cada indivíduo. Assim, para a realização da experiência, basta obter as  $n$  observações da *v.a.r.*  $Y$ .

Para enunciar convenientemente um resultado que assegura a consistência do estimador  $N$ - $W$ , consideremos as seguintes hipóteses:

$$(H_5) \int |K(s)| ds < \infty \text{ e } \lim_{|s| \rightarrow \infty} sK(s) = 0, \forall s \in \mathbb{R};$$

$$(H_6) E(Y^2) < \infty;$$

$$(H_7) h \rightarrow 0 \text{ e } nh \rightarrow +\infty.$$

**Proposição 2.2.** *Assumindo as hipóteses  $(H_5)$  a  $(H_7)$  acima listadas, tem-se*

$$\widehat{m}_{NW}(x) \xrightarrow{P} m(x), \quad (2.20)$$

para todo  $x$  ponto de continuidade de  $f, m$  e  $v$ , tal que  $f(x) > 0$ .

Podemos reescrever

$$\widehat{m}_{NW}(x) = \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i}{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i)} = \frac{\widehat{r}_n(x)}{\widehat{f}_n(x)}. \quad (2.21)$$

A demonstração da proposição 2.2 consistirá nas três etapas seguintes:

1. Provar que  $\widehat{r}_n(x) \xrightarrow{P} r(x) = m(x)f(x)$ ;
2. Provar que  $\widehat{f}_n(x) \xrightarrow{P} f(x)$ ;
3. Aplicar o corolário 5.7 (consequência dos teoremas 5.5 e 5.6) de Gonçalves e M. Lopes [?] para concluir que  $\frac{\widehat{r}_n(x)}{\widehat{f}_n(x)} \xrightarrow{P} m(x)$ . Tal corolário estabelece que, dado duas variáveis aleatórias reais  $Z$  e  $U$ , se existirem duas sucessões de variáveis aleatórias reais  $(Z_n)_{n \in \mathbb{N}}$  e  $(U_n)_{n \in \mathbb{N}}$  tais que  $Z_n \xrightarrow{P} Z$ ,  $U_n \xrightarrow{P} U$  e  $P(U = 0) = 0$ , então  $\frac{Z_n}{U_n} \xrightarrow{P} \frac{Z}{U}$ .

*Demonstração.* Esboçaremos a prova somente para a primeira etapa, uma vez que para a segunda as técnicas são as mesmas.

Por (1.4), tem-se

$$MSE(\widehat{r}_n(x)) = Var(\widehat{r}_n(x)) + (\mathbb{E}(\widehat{r}_n(x)) - r(x))^2. \quad (2.22)$$

Começemos por provar que o segundo termo do lado direito da equação anterior,  $(\mathbb{E}(\widehat{r}_n(x)) - r(x))^2$ , converge para 0. Ora,

$$\begin{aligned} |\mathbb{E}(\widehat{r}_n(x)) - r(x)| &\leq \int |K_h(s)| |r(x-s) - r(x)| ds \\ &\leq \int_{|s| \leq \delta} |K_h(s)| |r(x-s) - r(x)| ds \\ &\quad + \int_{|s| > \delta} |K_h(s)| |r(x-s)| ds \\ &\quad + \int_{|s| > \delta} |K_h(s)| |r(x)| ds \\ &= T_1 + T_2 + T_3. \end{aligned} \quad (2.23)$$

$$\begin{aligned} T_1 &= \int_{|s| \leq \delta} |K_h(s)| |r(x-s) - r(x)| ds \\ &\leq \sup_{|s| \leq \delta} |r(x-s) - r(x)| \int_{|s| \leq \delta} |K_h(s)| ds \\ &= \sup_{|s| \leq \delta} |r(x-s) - r(x)| \int |K(s)| ds \end{aligned} \quad (2.24)$$

quando  $h \rightarrow 0$ . Se fixarmos  $\delta$  arbitrariamente próximo de 0, pela continuidade de  $r$  e finitude de  $\int |K(s)| ds$ , a quantidade  $T_1$  será arbitrariamente pequena, isto é, relativamente próximo de 0.

$$\begin{aligned} T_2 &= \int_{|s| > \delta} |K_h(s)| |r(x-s)| ds \\ &\leq \frac{1}{\delta} \sup_{|s| > \frac{\delta}{h}} |sK(s)| \int_{|s| > \delta} |r(x-s)| ds \\ &\leq \frac{1}{\delta} \sup_{|s| > \frac{\delta}{h}} |sK(s)| \int |r(s)| ds. \end{aligned} \quad (2.25)$$

Uma vez que  $\delta$  está fixo, pela condição  $\lim_{|s| \rightarrow \infty} sK(s) = 0$ , concluímos que  $T_2 \rightarrow 0$  quando  $h \rightarrow 0$ .

$$T_3 = |r(x)| \int_{|s| > \frac{\delta}{h}} |K(s)| ds \rightarrow 0 \quad (2.26)$$

quando  $h \rightarrow 0$ , pois  $\int |K(s)| ds$  é finito e  $\delta$  está fixo.

Por (2.24), (2.25) e (2.26) concluímos que a quantidade  $|\mathbb{E}(\widehat{r}_n(x)) - r(x)|$  se torna tão pequena quanto se queira e, portanto, daí se conclui que converge para 0.

Para a variância, consideremos  $s^2(x) = E(Y^2|X = x)$ .

$$\begin{aligned}
 \text{Var}(\widehat{r}_n(x)) &= \frac{1}{n} \{E(K_h^2(x - X)Y^2) - E^2(K_h(x - X)Y)\} \\
 &= \frac{1}{n} \iint K_h^2(x - u)y^2 f(y|u)f(u)du dy \\
 &\quad - \frac{1}{n} \left( \iint K_h(x - u)yf(y|u)f(u) \right)^2 \\
 &= \frac{1}{n} \int K_h^2(x - u)s^2(u)f(u)du - \frac{1}{n} \left( \int K_h(x - u)r(u)du \right)^2 \\
 &= \frac{1}{n} \left\{ \frac{1}{h} \int K(s)s^2(x + sh)f(x + sh)ds - \left( \int K_h(s)r(x - s)ds \right)^2 \right\} \\
 &\approx \frac{1}{nh} s^2(x)f(x) \int K^2(s)ds, \tag{2.27}
 \end{aligned}$$

usando as mesmas técnicas de separação de integral que no caso anterior. Observemos que  $\text{Var}(\widehat{r}_n(x)) \rightarrow 0$ , desde que  $nh \rightarrow +\infty$ .

Concluimos então que  $MSE(\widehat{r}_n(x)) \rightarrow 0$  ( $MSE(\widehat{f}_n(x)) \rightarrow 0$ ) e, consequentemente,  $\widehat{r}_n(x) \xrightarrow{P} r(x)$  ( $\widehat{f}_n(x) \xrightarrow{P} f(x)$ ). Finalmente, pela etapa 3, concluimos que

$$\widehat{m}_{NW}(x) \xrightarrow{P} m(x),$$

conforme se pretendia. □

O resultado que acabamos de provar, que pode ser encontrado em Härdle [?], vem confirmar a poderosa estrutura do estimador  $N$ - $W$  e, mais ainda, impondo unicamente à curva (desconhecida)  $m$  que esta seja contínua e imputando estimativas pontuais (isto é, uma estimativa em cada ponto) à mesma curva, vem afirmar a sua maior flexibilidade, relativamente ao *regressograma*.

Outros resultados sobre a consistência deste estimador, alguns requerendo pelo menos a diferenciabilidade de  $m$ , podem ser encontrados em Härdle, Müller, Sperlich, and Werwatz [?].

Concluindo, temos a dizer que na secção anterior e nesta, introduzimos dois estimadores para a *função de regressão* que, conforme pudemos constatar, são ambos consistentes. O primeiro destes, o *regressograma*, sendo um estimador mais natural, com objectivo unicamente de introduzir a noção de estimação. Já o segundo,  $N$ - $W$ , pelas suas características (consistência, estimação pontual e interpolação das observações quando a sucessão  $h$  ( $h_1, \dots, h_d$ ) converge para 0 ( $0, \dots, 0$ )) será uma ferramenta de extrema importância para o trabalho que vamos desenvolver no próximo capítulo.

**Exemplo 2.2** (Aplicação do  $N-W$ ).

Consideremos o modelo

$$Y = \sqrt[3]{X} + \varepsilon \quad (2.28)$$

do exemplo 2.1 e, para uma comparação legítima entre a curva estimada construída à custa do *regressograma* e a curva estimada que iremos construir à custa do  $N-W$ , as mesmas 100 observações de  $X$ ,  $\varepsilon$  e portanto  $Y$ , isto é,  $\{Y_i\}_{i=1}^{100}$ , com  $Y_i = \sqrt[3]{X_i} + \varepsilon_i$ . Tais observações encontram-se dispostas no gráfico da Figura 2.1.

Neste exemplo, fixando  $h = 0.1$ , vamos construir duas curvas estimadas utilizando o estimador  $N-W$ : Uma delas usando o *kernel* uniforme e a outra o *kernel* gaussiano.

Os gráficos das figuras 2.3 e 2.4 vêm confirmar a maior regularidade das curvas estimadas usando o estimador  $N-W$ , relativamente às curvas estimadas usando o *regressograma* (ver o gráfico da Figura 2.2). Este facto é também claramente realçado pelo gráfico da Figura 2.6.

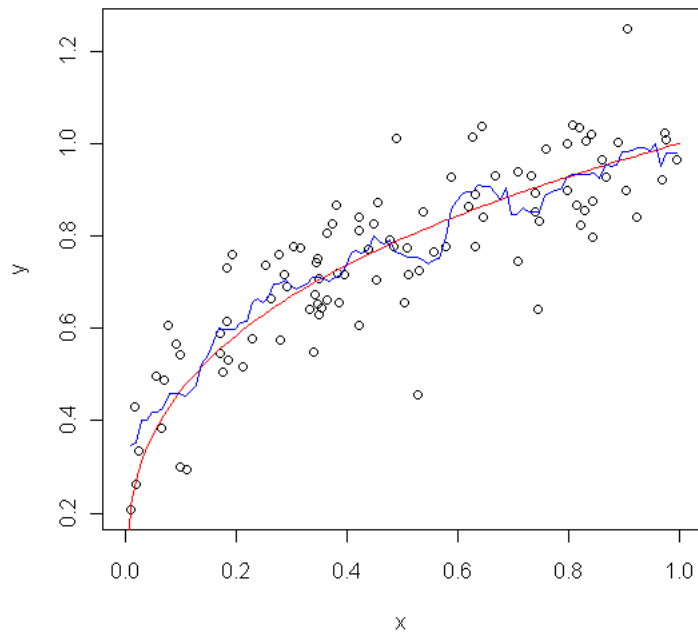


Figura 2.3: Representação gráfica da curva  $m$  e do  $N-W$  com o *kernel* uniforme. A curva  $m$  encontra-se a **vermelho** e o  $N-W$  a **azul**.

Relativamente ao estimador  $N-W$ , observando o gráfico da Figura 2.5, podemos ainda dizer que a escolha da função *kernel* a usar, uma vez que as duas curvas estimadas quase que se sobrepõe, não é muito relevante. Na verdade, prova-se que assintoticamente as estimativas de *kernel* para a *fdp* ( $\widehat{f}_n(x)$  e  $\widehat{g}_n(y, x)$ ), no nosso

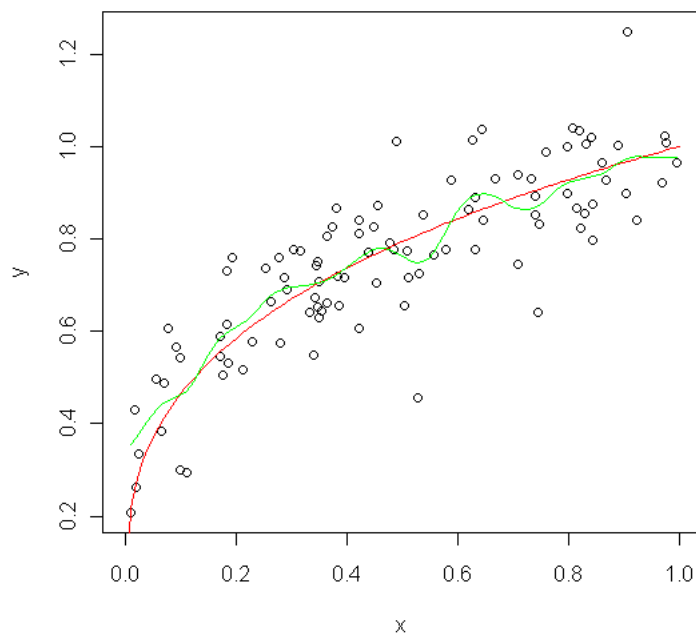


Figura 2.4: Representação gráfica da curva  $m$  e do  $N-W$  com o kernel gaussiano. A curva  $m$  encontra-se a **vermelho** e o  $N-W$  a **verde**.

caso), fixando o número de observações e a sucessão decrescente para 0,  $h$  ( $h$  e  $l$ , no nosso caso), não variam muito relativamente à função *kernel* usada, pelo que, conseqüentemente, o  $N-W$  ( $\widehat{m}_{NW}(x)$ , no nosso caso) não deverá variar muito. Estas provas podem ser encontradas em Härdle, Müller, Sperlich and Werwatz [?] ou Wand and Jones [?].



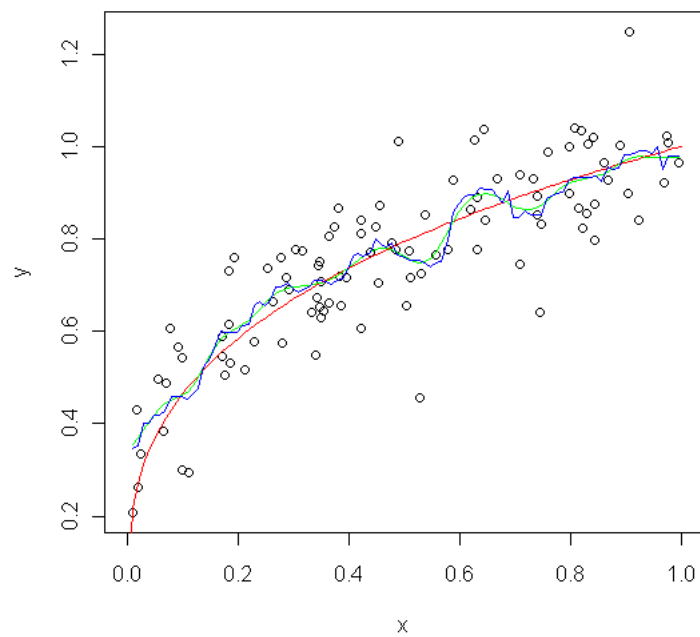


Figura 2.5: Representação gráfica da curva  $m$ , do N-W com o kernel uniforme e do N-W com kernel gaussiano. A curva  $m$  encontra-se a **vermelho**, o N-W com kernel uniforme a **azul** e o N-W com kernel gaussiano a **verde**.

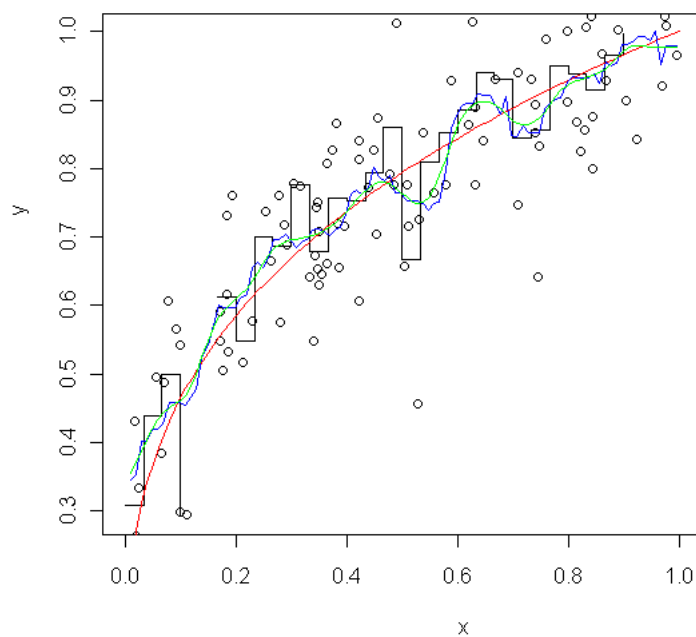


Figura 2.6: Representação gráfica da curva  $m$ , do regressograma, do N-W usando o kernel uniforme e do N-W usando o kernel gaussiano. A curva  $m$  encontra-se a **vermelho**, o regressograma a **preto**, o N-W usando kernel uniforme a **azul** e o N-W usando kernel gaussiano a **verde**.

## Capítulo 3

# Estimação com observações em falta

Todos que de uma forma ou outra lidam com a estatística estão susceptíveis a confrontar ou já confrontaram com o problema de estimação com observações em falta. Este tipo de problema é tão comum que existem vários livros que lhe são inteiramente dedicados (consultar, por exemplo, Allison [?] e Little and Rubin [?]). De facto, elevados custos para ter acesso a algumas informações ou parte destas, perda de informação causada por factores incontroláveis e falha dos investigadores na recolha de informação, são algumas das razões que podem levar a que nem sempre as observações estejam todas disponíveis. Nestes casos, não sendo sempre possível aplicar directamente os procedimentos habituais (nas inferências), há que adaptá-los.

Neste capítulo vamos abordar um problema particular de estimação com observações em falta na amostra. Concretamente, abordaremos o problema de estimação da média de uma variável aleatória, que depende de outra (ou outras) através de um modelo de regressão, quando algumas das observações que compõe a amostra se perdem, nomeadamente, quando, na amostra que permite o tratamento do problema de regressão, algumas das ordenadas não são conhecidas, mas sim apenas o valor para variável explicativa.

### 3.1. Introdução

Consideremos uma sucessão de observações  $\{(Y_i, X_i)\}_{i=1}^n$  em que cada par  $(Y_i, X_i)$  é *i.i.d.* com  $(Y, X)$ , onde  $Y$  é a *v.a.r.* que se pretende explicar e  $X$  é agora assumido como sendo o vector de  $d$  *v.a.r.* explicativas. Pois, no capítulo anterior consideramo-lo unidimensional com o objectivo único de permitir uma abordagem mais simplificada dos estimadores que ali foram introduzidos, pelo que agora retomamos ao caso genérico,  $X \in \mathbb{R}^d$ . Designemo-lo por vector aleatório real (*ve.a.r.*).

De agora em diante assumamos a notação  $(X, Y)$  em vez de  $(Y, X)$ . Com isto pretendemos ter uma noção espacial de  $Y$  e  $X$  como se de ordenada e abcissa (no

referencial plano de coordenadas cartesianas), respectivamente, se tratassem e consequentemente uma melhor compreensão do problema que abordaremos. Isto tudo quando olhamos para  $X$  como unidimensional, obviamente.

Passemos então à introdução do problema. Neste sentido, suponhamos que todas as  $n$  observações do *ve.a.r.*  $X$  estão disponíveis, mas, por alguma razão, algumas das observações da *ve.a.r.*  $Y$  perderam-se. Mais precisamente, suponhamos que observamos  $n$  vezes o par  $(X, Y)$ , pelo que obtemos a sucessão de observações  $\{(X_i, Y_i)\}_{i=1}^n$ , mas, a certa altura constatamos que alguns dos  $Y_i$ 's,  $i = 1, \dots, n$ , perderam-se. Suponhamos ainda que estes perderam-se de forma aleatória, isto é, antes de observar o par  $(X, Y)$ , não sabemos quais serão os índices  $i$ 's para os quais os  $Y_i$ 's estarão em falta.

Assim, a amostra com que ficamos contém todas as  $n$  observações do vector aleatória  $X$ , mas algumas das observações da variável aleatória  $Y$  encontram-se *aleatoriamente em falta* e, portanto, podemos dizer que a nossa amostra final é a sucessão de observações  $\{(X_i, Y_i)\}_{i=1}^n$ , com alguns dos  $Y_i$ 's *aleatoriamente em falta*.

Para descrever convenientemente o problema e por forma a permitir o tratamento matemático, consideremos uma sucessão de observações  $\{\delta_i\}_{i=1}^n$  da variável de Bernoulli  $\delta$ , tal que

$$\delta_i = \begin{cases} 1 & \text{se } Y_i \text{ está disponível} \\ 0 & \text{se } Y_i \text{ não está disponível,} \end{cases} \quad (3.1)$$

pelo que na prática, obtemos uma sucessão de observações incompletas

$$\{(X_i, Y_i, \delta_i)\}_{i=1}^n, \quad (3.2)$$

onde  $(X_i, Y_i, \delta_i)$ ,  $i = 1, \dots, n$ , é *i.i.d.* com  $(X, Y, \delta)$ .

Neste trabalho estamos interessados na inferência sobre a média da v.a.r.  $Y$ , que denotaremos por  $\theta$ , quando há observações da mesma *aleatoriamente em falta* na sucessão  $\{(X_i, Y_i)\}_{i=1}^n$ . Neste sentido, vamos, nas duas próximas secções, introduzir dois estimadores que, como veremos, permitem ambos imputar consistentes estimativas às observações em falta e estimar  $\theta$  como se as  $n$  observações de  $Y$  estivessem todas disponíveis. Tais estimadores foram introduzidos por Cheng [?] e Wang, Linton, and Härdle [?] usando os *MRNP* e *MRPL*, respectivamente, para imputar consistentes estimativas às observações em falta, conforme poderemos constatar. Relativamente ao *MRPL*, deixamos a sua introdução e abordagem para mais adiante (subsecção 3.3.1), sendo que a abordagem será feita no contexto do problema em estudo.

Obviamente, uma alternativa à abordagem referida no parágrafo anterior seria ignorar as observações que se encontram em falta e, aplicando as técnicas convencionais, estimar  $\theta$  usando somente as observações disponíveis. Contudo, o nosso objectivo aqui é tirar partido de que cada  $Y_i$  em falta fora observado e que este está relacionado com  $X_i$  através de um modelo de regressão.

Neste trabalho, de modo semelhante a Cheng [?] e Wang, Linton, and Härdle [?], vamos convenientemente abordar a assumpção “obsevações da v.a.r.  $Y$  aleatoriamente em falta” assumindo que  $\delta$  e  $Y$  são condicionalmente independentes dado o vector aleatório  $X$ . Isto é,

$$P(\delta = 1|Y, X) = P(\delta = 1|X) \quad (3.3)$$

(de modo equivalente,  $E(\delta|Y, X) = E(\delta|X)$ ). Denotemos por  $p(X)$  a igualdade anterior, isto é,  $p(X) = P(\delta = 1|X)$ . Para uma discussão acerca da caracterização do sistema de amostras com observações em falta, recomendamos Little and Rubin [?].

### 3.2. Estimador de Cheng para a média - Modelo não paramétrico

No capítulo 1 introduzimos alguns modelos para calcular a regressão de  $Y$  sobre  $X$ , entre os quais o *MRNP* e o *MRPL*. Tais modelos, conforme veremos, constituem uma óptima ferramenta para inferir sobre as observações em falta.

Consideremos uma sucessão de observações  $\{(X_i, Y_i)\}_{i=1}^n$ , onde cada par  $(X_i, Y_i)$  é *i.i.d* com  $(X, Y)$ . O *MRNP* permite-nos escrever a seguinte relação:

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \dots, n, \quad (3.4)$$

onde  $X_i = (X_{i1}, \dots, X_{id})^T$ ,  $i = 1, \dots, n$ ,  $m$  é a *função de regressão* e os erros  $\varepsilon_1, \dots, \varepsilon_n$  são *i.i.d.* com a v.a.r.  $\varepsilon$  de média nula e variância finita. Este modelo, conforme tínhamos visto, não impõe nenhuma restrição sobre a forma funcional de  $m$ .

Cheng [?] abordou eficientemente o problema introduzido na secção anterior, considerando as observações de  $Y$  *aleatoriamente em falta* no modelo (3.4). Pois, abordando-o desta forma, estará em condições de utilizar as técnicas de regressão, mais concretamente os estimadores para a *função de regressão*, para imputar consistentes estimativas às observações em falta. De facto, por exemplo, o estimador *N-W* associado à sucessão  $\{(X_i, Y_i)\}_{i=1}^n$ , conforme podemos constatar na secção 2.2,

quando aplicado nos pontos do conjunto  $\{X_i, i = 1, \dots, n\}$ , com  $n$  fixo, obtém-se o conjunto de pontos  $\{Y_i, i = 1, \dots, n\}$  se  $h$  for escolhido suficientemente próximo de 0. Isto é,  $\widehat{m}_{NW}(X_i) \rightarrow Y_i$  à medida que formos tornando  $h$  cada vez mais pequeno. Assim, aproveitando-se desta característica do  $N$ - $W$ , poder-se-á “recuperar” todos os  $Y_i$ 's que se encontram em falta e, conseqüentemente, estimar  $\theta$  como se as  $n$  observações da *v.a.r*  $Y$  estivessem completas. De facto, foi este o procedimento adoptado por Cheng [?] na sua abordagem ao problema.

Ora, Cheng [?] usou o  $N$ - $W$  associado ao conjunto de observações  $\{(X_i, Y_i), i = 1, \dots, n, \delta_i = 1\}$  para encontrar uma curva aproximada para  $m$  e, para “recuperar” o conjunto de pontos (em falta)  $\{Y_i, i = 1, \dots, n, \delta_i = 0\}$ , aplicou-o ao conjunto  $\{X_i, i = 1, \dots, n, \delta_i = 0\}$  sob a condição  $h \rightarrow 0$ . Isto é, imputou uma estimativa  $\widehat{Y}_i = \widehat{m}_{NW}(X_i)$  a cada  $Y_i$  tal que  $\delta_i = 0, i = 1, \dots, n$ , pelo que deste modo pôde estimar  $\theta$  como se as  $n$  observações de  $Y$  estivessem completas. Entretanto, construiu o estimador

$$\begin{aligned} \widehat{\theta}_{ch} &= \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i Y_i + (1 - \delta_i) \widehat{Y}_i \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i Y_i + (1 - \delta_i) \widehat{m}_{NW}(X_i) \right\} \end{aligned} \quad (3.5)$$

para  $\theta$  e provou a *normalidade assintótica* para a versão aparada

$$\widehat{\theta}_{ch}^a = \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i Y_i + (1 - \delta_i) \widehat{m}_{NW}^a(X_i) \right\}, \quad (3.6)$$

onde

$$\widehat{m}_{NW}^a(x) = \widehat{m}_{NW}(x) I_{\left\{ \sum_{j=1}^n K_h(x - X_j) \delta_j \geq \frac{c}{h} \log(n) \right\}},$$

em que  $I_{\{\cdot\}}$  é a função indicatriz e a constante  $c$  é o coeficiente de apáero. A prova da *normalidade assintótica* é esboçada para a versão aparada, pelo facto de  $\theta$  ser uma média global, mas, segundo Cheng [?], é também possível prova-la para a versão original  $\widehat{\theta}_{ch}$  (ver nota 3.1).

Definamos  $q(x) = p(x)f(x)$ ,  $\widehat{q}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \delta_i$  e  $\sigma^2(X) = \text{Var}(Y|X)$ .

O teorema que seguidamente enunciaremos assegura a *normalidade assintótica* de  $\widehat{\theta}_{ch}^a$  para o caso  $X$  unidimensional (para o caso geral, consultar Cheng [?]). Antes de enunciá-lo, consideremos as seguintes hipóteses:

$$(H_8) \quad E(Y^2) < \infty;$$

$$(H_9) \quad E\left(\frac{\sigma^2(X)}{p(X)}\right) < \infty \text{ e } E(p(X)) = p(\delta = 1) = p \in ]0, 1];$$

(H<sub>10</sub>)  $q$  e  $m$  admitem derivadas de segunda ordem tais que

$$E\left\{(\sigma(X) + m(X)) \frac{q''(X)}{\sqrt{q(X)}} (1 - p(X))\right\} < \infty$$

e

$$E\left\{(m(X)(q''(X))^2 + m'(X)q'(X)) \frac{1}{q(X)} (1 - p(X))\right\} < \infty;$$

(H<sub>11</sub>) A função *kernel* é uma densidade de probabilidade limitada e simétrica sobre  $\mathbb{R}$  tal que  $\int s^2 K(s) ds < \infty$ ;

(H<sub>12</sub>)  $h$  é tal que  $nh^2 \rightarrow +\infty$  e  $\frac{nh^3}{\log(n)} \rightarrow 0$ .

**Teorema 3.1** (Cheng [?]). *Assumindo as hipóteses (H<sub>8</sub>) a (H<sub>12</sub>) acima listadas e (3.3) para o conjunto de dados incompletos (3.2), tem-se*

$$\sqrt{n} \left( \hat{\theta}_{ch}^a - \theta \right) \xrightarrow{L} \mathcal{N} \left( 0, \sigma^2 \right),$$

onde  $\sigma^2 = E \left( \frac{\sigma^2(X)}{p(X)} \right) + Var(m(X))$  e  $\mathcal{N} \left( 0, \sigma^2 \right)$  a lei normal de média nula e variância  $\sigma^2$ .

**Nota 3.1.** Segundo Cheng [?], o teorema anterior assegura a *normalidade assintótica* para a versão original  $\hat{\theta}_{ch}$  se, para além das hipóteses (H<sub>8</sub>) a (H<sub>12</sub>) acima listadas, assumirmos  $X$  com suporte compacto ou as seguintes condições:

- $\forall x$  nas caudas da distribuição de  $X$ , existe  $\alpha > 0$  tal que  $p(x) \geq \alpha$ ;
- $X$  decresce exponencialmente;
- $\forall X$ ,  $m(x)$  é limitado superiormente pelo valor duma função polinomial em  $x$ .

*Demonstração.* Começemos por reescrever  $\hat{\theta}_{ch}^a$  na forma

$$\begin{aligned} \hat{\theta}_{ch}^a - \theta &= \frac{1}{n} \sum_{i=1}^n (m(X_i) - \theta) + \frac{1}{n} \sum_{i=1}^n \delta_i (Y_i - m(X_i)) \\ &\quad + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left( \widehat{m}_{NW}^a(X_i) - m(X_i) \right) \\ &= R_n + S_n + T_n. \end{aligned} \tag{3.7}$$

Tendo em conta que  $m(X)$  é uma função mensurável da v.a.r.  $X$ ,  $m(X)$  e  $\theta$  são não correlacionados (pois,  $Cov(m(X), \theta) = E(m(X)\theta) - E(m(X))E(\theta) = 0$ ), as observações de  $X$ ,  $Y$  e  $\delta$  são *i.i.d.* e  $E(\delta|Y, X) = E(\delta|X)$ , calculemos  $E(\sqrt{n}R_n)$ ,  $E(\sqrt{n}S_n)$ ,

$Var(\sqrt{n}R_n)$ ,  $Var(\sqrt{n}S_n)$  e  $Cov(\sqrt{n}R_n, \sqrt{n}S_n)$ , onde  $R_n = \frac{1}{n} \sum_{i=1}^n (m(X_i) - \theta)$  e  $S_n = \frac{1}{n} \sum_{i=1}^n \delta_i (Y_i - m(X_i))$ . Para as esperanças, tem-se

$$E(\sqrt{n}R_n) = \sqrt{n} E(m(X) - \theta) = 0 \quad (3.8)$$

e

$$\begin{aligned} E(\sqrt{n}S_n) &= \sqrt{n} E[\delta Y - \delta m(X)] \\ &= \sqrt{n} E[E(\delta Y|X) - m(X) E(\delta|X)] \\ &= 0. \end{aligned} \quad (3.9)$$

Para as variâncias, tem-se

$$Var(\sqrt{n}R_n) = Var(m(X)) \quad (3.10)$$

e

$$\begin{aligned} Var(\sqrt{n}S_n) &= Var[\delta(Y - m(X))] \\ &= E[\delta(Y - m(X))^2] \\ &= E[E(\delta Y^2|X) - 2m(X) E(\delta Y|X) + E(\delta m^2(X)|X)] \\ &= E(E(\delta|X) Var(Y|X)) \\ &= E(p(X)\sigma^2(X)). \end{aligned} \quad (3.11)$$

Para a covariância, tem-se

$$\begin{aligned} Cov(\sqrt{n}R_n, \sqrt{n}S_n) &= E[\delta(m(X) - \theta)(Y - m(X))] \\ &= E[m(X) E(\delta Y|X)] - E[m^2(X) E(\delta|X)] \\ &\quad - E[\theta E(\delta Y|X)] + E[\theta m(X) E(\delta|X)] \\ &= 0. \end{aligned} \quad (3.12)$$

Pelas equações (3.8), (3.9), (3.10), (3.11) e (3.12) e pelo *teorema do limite central* concluímos que  $\sqrt{n}(R_n + S_n)$  converge para a *lei normal* de média nula e variância  $Var(m(X)) + E(p(X)\sigma^2(X))$ .

Para o último termo do lado direito da equação (3.7),

$$T_n = \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left( \widehat{m}_{NW}^{\alpha}(X_i) - m(X_i) \right),$$

seja  $T_n = T' + (T_n - T')$ , onde  $T' = E[T_n | (X_i, Y_i), i = 1, \dots, n, \delta_i = 1]$ . Provemos que  $(T_n - T')$  é um  $o(1)$ , provando que  $T_n$  converge em média quadrática para  $T'$ .



De facto, observando que  $E(E[(T_n - T')|(X_i, Y_i), i = 1, \dots, n, \delta_i = 1]) = 0$ , tem-se

$$\begin{aligned} E(T_n - T')^2 &= E(E[(T_n - T')^2|(X_i, Y_i), i = 1, \dots, n, \delta_i = 1]) \\ &= E(Var[T_n - T'|(X_i, Y_i), i = 1, \dots, n, \delta_i = 1]) \\ &= \Lambda_{n1} + \Lambda_{n2} + \Lambda_{n3}, \end{aligned} \quad (3.13)$$

onde

$$\begin{aligned} \Lambda_{n1} &= \frac{1}{n} E \left( Var \left\{ (1 - \delta_i)(\widehat{m}_{NW}^a(X_i) - m(X_i)) \middle| (X_i, Y_i), i = 1, \dots, n, \delta_i = 1 \right\} \right) \\ &= \frac{1}{n} E \left( E \left[ (1 - \delta_i)(\widehat{m}_{NW}^a(X_i) - m(X_i))^2 \middle| (X_i, Y_i), i = 1, \dots, n, \delta_i = 1 \right] \right) \\ &\quad - \frac{1}{n} E(T')^2 \\ &= \frac{1}{n} E \left[ (1 - \delta_i) (\widehat{m}_{NW}^a(X_i) - m(X_i))^2 \right] - \frac{1}{n} E(T')^2 \end{aligned} \quad (3.14)$$

e

$$\Lambda_{n2} = E[Var \{ \phi(\delta_i, X_i, Y_i) | (X_i, Y_i), i = 1, \dots, n, \delta_i = 1 \}] = 0, \quad (3.15)$$

com  $\phi(\delta_i, X_i, Y_i) = E \left[ (1 - \delta_i) (\widehat{m}_{NW}^a(X_i) - m(X_i)) \middle| (X_i, Y_i), i = 1, \dots, n, \delta_i = 1 \right]$ . O resultado (3.14) obtém-se somando e subtraindo  $\widehat{m}_{NW}^a(X_i)$  ao termo  $\widehat{m}_{NW}^a(X_i) - m(X_i)$  e usando o facto de  $K$  ser uma *fdp* limitada.

Para o termo  $\Lambda_{n3}$ , seja  $\chi(\delta_i, X_i) = (1 - \delta_i) (\widehat{m}_{NW}^a(X_i) - m(X_i))$ ,

$$\Omega_i = \{(X_i, Y_i), i = 1, \dots, n, \delta_i = 1\}$$

e  $\phi$  definido acima. Assim,

$$\begin{aligned} \Lambda_{n3} &= -2 E \left[ Cov \left\{ \left[ \frac{1}{n} \sum_{i=1}^n \chi(\delta_i, X_i), \phi(\delta_i, X_i, Y_i) \right] \middle| \Omega_i \right\} \right] \\ &= -2 E \left\{ E \left[ \phi(\delta_i, X_i, Y_i) \frac{1}{n} \sum_{i=1}^n \chi(\delta_i, X_i) \middle| \Omega_i \right] \right\} \\ &\quad + 2 E \left\{ E \left[ \frac{1}{n} \sum_{i=1}^n \chi(\delta_i, X_i) \middle| \Omega_i \right] E[\phi(\delta_i, X_i, Y_i) | \Omega_i] \right\} \\ &= -2 E \{ \phi(\delta_i, X_i, Y_i) E[\chi(\delta_i, X_i) | \Omega_i] \} \\ &\quad + 2 E \{ E[\chi(\delta_i, X_i) | \Omega_i] \phi(\delta_i, X_i, Y_i) \} \\ &= 0. \end{aligned} \quad (3.16)$$

Por (3.13), (3.14), (3.15) e (3.16) concluímos que

$$E(T_n - T')^2 = \frac{1}{n} E \left[ (1 - \delta_i) (\widehat{m}_{NW}^a(X_i) - m(X_i))^2 \right] - \frac{1}{n} E(T')^2 \quad (3.17)$$

e, pelo facto de  $E(T')^2 \geq 0$  e  $E\left[\delta_i (\widehat{m}_{NW}(X_i) - m(X_i))^2\right] \geq 0$ , concluímos que

$$n E(T_n - T')^2 \leq E\left[(\widehat{m}_{NW}(X_i) - m(X_i))^2\right]. \quad (3.18)$$

Pelo teorema 2 de Spiegelman and Sacks [?],  $E\left[(\widehat{m}_{NW}(X_i) - m(X_i))^2\right] \rightarrow 0$  quando  $n \rightarrow +\infty$ , pelo que, conseqüentemente, também  $E(T_n - T')^2 \rightarrow 0$  quando  $n \rightarrow +\infty$ .

Resta-nos, portanto, concluir a *normalidade assintótica* de  $\sqrt{n}(R_n + S_n + T')$ . Observemos que a condição  $\{\widehat{q}(x) \geq \frac{c}{nh} \log(n)\}$  é equivalente a

$$\left\{q(x) \geq \frac{c}{nh} \log(n) - (\widehat{q}(x) - q(x))\right\},$$

onde o termo  $(\widehat{q}(x) - q(x))$  converge em probabilidade para 0 (pela proposição 2.2). Portanto, existem constantes positivas  $c$ ,  $c_1$  e  $c_2$  tais que  $\{\widehat{q}(x) \geq \frac{c}{nh} \log(n)\}$  reduz-se a  $\{(\widehat{q}(x) \geq c_1)\}$  sempre que  $\{(q(x) \geq c_2)\}$ . Isto implica que  $\sqrt{n}T'$  e  $\sqrt{n}T' I_{\{A_n\}}$  são assintoticamente equivalentes em probabilidade, onde  $A_n = \{\widehat{q}(x) \geq c_1\} \cap \{q(x) \geq c_2\}$ . Portanto,  $\sqrt{n}T' I_{\{A_n\}}$  é assintoticamente equivalente a  $\sqrt{n}U_n$ , onde

$$U_n = \frac{1}{n} \sum_{i=1}^n \delta_i \{Y_i - m(X_i)\} E\left[K_h(X - X_i) \left(\frac{1-p(X)}{q(X)}\right) I_{\{A_n\}} \middle| X_i\right],$$

pelo que, em vez de  $\sqrt{n}(R_n + S_n + T')$ , vamos, de modo equivalente, concluir a *normalidade assintótica* de  $\sqrt{n}(R_n + S_n + U_n)$ .

Vejamos que  $\sqrt{n}T' I_{\{A_n\}}$  é de facto assintoticamente equivalente a  $\sqrt{n}U_n$ . Com efeito,

$$\begin{aligned} \sqrt{n}T' I_{\{A_n\}} &= E\left[\frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left(\widehat{m}_{NW}(X_i) - m(X_i)\right) \middle| \Omega_i\right] I_{\{A_n\}} \\ &= E\left[\frac{1}{n^2} \sum_{i=1}^n (1 - \delta_i) \sum_{j=1}^n K_h(X_i - X_j) \frac{\delta_j Y_j}{\widehat{q}(X_i)} I_{\{\widehat{q}(X_i) \geq \frac{c \log(n)}{hn}\}} \middle| \Omega_i\right] I_{\{A_n\}} \\ &\quad - E\left[\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (1 - \delta_i) m(X_i) K_h(X_i - X_j) \frac{\delta_j}{\widehat{q}(X_i)} \middle| \Omega_i\right] I_{\{A_n\}} \\ &= \frac{1}{n} \sum_{j=1}^n \delta_j \{Y_j - m(X_j)\} E\left[K_h(X - X_j) \frac{(1-p(X))}{q(X)} I_{\{A_n\}} \middle| X_j\right] \\ &\quad + o_p(1), \end{aligned} \quad (3.19)$$

tendo em conta a simetria de  $K$ .

Notemos que  $E(\sqrt{n}U_n) = 0$  e

$$\begin{aligned} Var(\sqrt{n}U_n) &= E\left[\left(\frac{1-p(X)}{q(X)}\right)^2 E^2\{K_h(X - X_i) I_{\{A_n\}} \middle| X_i\} p(X) \sigma^2(X)\right] \\ &= E\left[\frac{\{1-p(X)\}^2}{p(X)} \sigma^2(X)\right], \end{aligned} \quad (3.20)$$

tendo em conta a igualdade  $q(x) = p(x)f(x)$ , pelo que  $\sqrt{n}U_n$  segue assintoticamente a *lei normal* de média nula e variância igual a  $E \left[ \frac{\{1-p(X)\}^2}{p(X)} \sigma^2(X) \right]$ . Notemos ainda que

$$\begin{aligned} Cov(\sqrt{n}R_n, \sqrt{n}U_n) &= Cov(\sqrt{n}R_n, \sqrt{n}S_n) \\ &= 0 \end{aligned} \quad (3.21)$$

e, tendo em conta a igualdade  $q(x) = p(x)f(x)$ ,

$$\begin{aligned} Cov(\sqrt{n}S_n, \sqrt{n}U_n) &= E \left[ E \left\{ K_h(X - X_i) \left( \frac{1-p(X)}{q(X)} \right) I_{\{A_n\}} \middle| X_i \right\} p(X) \sigma^2(X) \right] \\ &= E \left[ (1-p(X)) \sigma^2(X) \right]. \end{aligned} \quad (3.22)$$

Finalmente, reagrupando os resultados obtidos, concluímos que

$$E(\sqrt{n}(R_n + S_n + U_n)) = 0,$$

$$\begin{aligned} Var(\sqrt{n}(R_n + S_n + U_n)) &= E[p(X)\sigma^2(X)] + Var(m(X)) \\ &\quad + E \left[ \frac{\{1-p(X)\}^2}{p(X)} \sigma^2(X) \right] + 2E[(1-p(X))\sigma^2(X)] \\ &= E \left[ \frac{\sigma^2(X)}{p(X)} \right] + Var(m(X)) \end{aligned} \quad (3.23)$$

e, conseqüentemente,  $\sqrt{n}(R_n + S_n + U_n)$  converge para a *lei normal* de média nula e variância igual a  $E \left[ \frac{\sigma^2(X)}{p(X)} \right] + Var(m(X))$ , conforme o estabelecido pelo teorema. Esta prova pode ser encontrada em Cheng [?].  $\square$

Embora consistente, conforme o teorema anterior, o estimador  $\hat{\theta}_{ch}$  pode não ter um bom desempenho quando a dimensão de  $X$  for grande, por causa da já referida *maldição da dimensionalidade* (ver secção 1.2). Na próxima secção veremos uma abordagem alternativa para este caso.

Podemos também ser do nosso interesse introduzir um estimador para a variância assintótica,  $\sigma^2$ , podendo este ser uma ferramenta útil nas inferências sobre o intervalo de confiança e teste de hipóteses para  $\theta$ . Assim, consideremos o seguinte estimador para  $\sigma^2$ :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\hat{\sigma}^2(X_i)}{\hat{P}(X_i)} + \widehat{m}_{NW}^2(X_i) \right\} - \left( \frac{1}{n} \sum_{i=1}^n \widehat{m}_{NW}(X_i) \right)^2, \quad (3.24)$$

onde

$$\hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n \left( \frac{K_h(x - X_i) \delta_i}{\frac{1}{n} \sum_{j=1}^n K_h(x - X_j) \delta_j} \right) Y_i^2 - \widehat{m}_{NW}^2(x)$$

e

$$\widehat{P}(x) = \frac{1}{n} \sum_{i=1}^n \left( \frac{K_h(x - X_i)}{\frac{1}{n} \sum_{j=1}^n K_h(x - X_j)} \right) \delta_i.$$

Embora afirmando ser consistente, Cheng [?] introduziu este estimador sem apresentar nenhum resultado. Contudo, neste trabalho, vamos propor e demonstrar um resultado que garante a sua consistência.

**Proposição 3.2.** *Nas condições do teorema anterior*

$$\widehat{\sigma}^2 \xrightarrow{P} \sigma^2. \quad (3.25)$$

*Demonstração.* Nas condições do teorema 3.1 e pela proposição 2.2,  $\widehat{\sigma}^2(x) \xrightarrow{P} E(Y^2|X = x) - E^2(Y|X = x) = Var(Y|X = x)$ ,  $\widehat{P}(x) \xrightarrow{P} E(\delta|X = x) = P(x)$  e, conseqüentemente,  $\frac{\widehat{\sigma}^2(x)}{\widehat{P}(x)} \xrightarrow{P} \frac{\sigma^2(x)}{P(x)}$ . O termo

$$\frac{1}{n} \sum_{i=1}^n \widehat{m}_{NW}^2(X_i) - \left( \frac{1}{n} \sum_{i=1}^n \widehat{m}_{NW}(X_i) \right)^2$$

converge em probabilidade para  $\frac{1}{n} \sum_{i=1}^n m^2(X_i) - \left( \frac{1}{n} \sum_{i=1}^n m(X_i) \right)^2 = L_n$ , pela proposição 2.2. Pela *lei dos grandes números*, os termos  $\frac{1}{n} \sum_{i=1}^n \frac{\sigma^2(X_i)}{P(X_i)}$  e  $L_n$  convergem em probabilidade para  $E\left(\frac{\sigma^2(x)}{P(x)}\right)$  e  $E(m^2(x)) - E^2(m(x)) = Var(m(x))$ , respectivamente. Finalmente, reagrupando os termos, concluímos a demonstração.

Os resultados sobre a *lei dos grandes números* podem ser encontrados em Gonçalves e M. Lopes [?] e Monfort [?], entre outras referências.  $\square$

O estimador  $\widehat{\theta}_{ch}$ , pela sua consistência, pode constituir uma ferramenta valiosa nas inferências sobre o intervalo de confiança e teste de hipóteses para  $\theta$ . Com efeito, pelo teorema 3.1,  $\frac{\widehat{\theta}_{ch} - \theta}{\sigma \sqrt{1/n}}$  converge para a *lei normal standard* (média nula e variância 1) e, sendo  $\widehat{\sigma}^2$  um estimador consistente de  $\sigma^2$  conforme a proposição anterior, o intervalo  $\left[ \widehat{\theta}_{ch} - \sqrt{\frac{\widehat{\sigma}^2}{n}} s_{\frac{\alpha}{2}}, \widehat{\theta}_{ch} + \sqrt{\frac{\widehat{\sigma}^2}{n}} s_{\frac{\alpha}{2}} \right]$  é um intervalo de confiança para  $\theta$  ao nível  $1 - \alpha$ , onde  $s_{\frac{\alpha}{2}}$  é o quantil de ordem  $\frac{\alpha}{2}$  da *lei normal standard*. Relativamente a teste de hipóteses, pela natureza da sua formulação matemática, este merece um destaque que a estrutura deste trabalho não permite. Detalhes acerca deste tema podem ser encontrados em Lehmann [?].

### 3.3. Estimador para a média - Modelo parcialmente linear

Nesta secção, como objectivo principal deste trabalho, vamos introduzir e analisar um novo estimador para  $\theta$ . Esse estimador, em vez de considerar observações de  $Y$  aleatoriamente em falta no *MRNP* como Cheng [?], considera-os no *MRPL*.

Para reformular o problema que foi introduzido na secção 3.1 e estudado na secção anterior, adaptando-o às características do *MRPL*, separemos o ve.a.r.  $X \in \mathbb{R}^d$  em  $X^* \in \mathbb{R}^p$  e  $T \in \mathbb{R}^{d^*}$ , onde  $p + d^* = d$ , ou seja,  $X = (X^*, T)$ , consideremos a sucessão de observações  $\{\delta_i\}_{i=1}^n$  da variável de Bernoulli  $\delta$ , definida na secção anterior, e uma sucessão de observações  $\{(X_i^*, T_i, Y_i)\}_{i=1}^n$ , onde  $(X_i^*, T_i, Y_i)$  é *i.i.d.* com  $(X^*, T, Y)$  e todos os  $n$  pares  $(X_i^*, T_i)$  estão disponíveis mas há observações da v.a.r.  $Y$  aleatoriamente em falta. Deste modo, obtemos uma sucessão de observações incompletas

$$\{(X_i^*, T_i, Y_i, \delta_i)\}_{i=1}^n, \quad (3.26)$$

onde  $(X_i^*, T_i, Y_i, \delta_i)$ ,  $i = 1, \dots, n$ , é *i.i.d.* com  $(X^*, T, Y, \delta)$ . Aqui, a condição definida por (3.3) assume a forma

$$P(\delta = 1|Y, X^*, T) = P(\delta = 1|X^*, T) = E(\delta|X^*, T). \quad (3.27)$$

Nesta secção, à semelhança de Cheng [?], estamos interessados na estimação da média de  $Y$ ,  $\theta$ , assumindo (3.27). Em vez do *MRNP* (3.4) proposto por Cheng [?], estimaremos  $\theta$  assumindo (3.27) no *MRPL*

$$Y_i = X_i^{*T} \beta + m^*(T_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.28)$$

onde  $X_i = (X_{i1}, \dots, X_{ip})^T$ ,  $T_i = (T_{i1}, \dots, T_{id^*})$ ,  $m^*$  é uma função desconhecida,  $\beta = (\beta_1, \dots, \beta_p)^T$  é um vector de parâmetros desconhecidos e os erros  $\varepsilon_1, \dots, \varepsilon_n$  são *i.i.d.* com a v.a.r.  $\varepsilon$  de média nula e variância finita. O *MRPL* pode ser interpretado como soma de uma parte puramente paramétrica,  $X^{*T} \beta = E(Y|X^*)$ , e outra puramente não paramétrica,  $m^*(T) = E(Y|T)$ . Sem ambiguidade, à função desconhecida  $m^*$  denotaremos por  $m$ .

Conforme havíamos referido no capítulo 1, sem surpresa nenhuma, estimar  $\beta$  e  $m$  envolve técnicas de *regressão paramétrica* e *não paramétrica*.

### 3.3.1. Estimação da função de regressão

Introduzido por Engle, Granger, Rice and Weiss [?] para estudar o efeito do tempo (temperatura) sobre o consumo da electricidade, o *MRPL* é na maioria das vezes preferível ao modelo completamente *não paramétrico* por causa da já mencionada *maldição da dimensionalidade* (ver secção 1.2). Obviamente, este modelo pressupõe que algumas das variáveis (marginais) explicativas em  $X$  influenciam linearmente a variável a ser explicada  $Y$ .

**Observação 3.1.** Notemos que aqui a *função de regressão* assume a forma

$$E(Y|X^*, T) = X^{*T}\beta + m(T),$$

com  $m(T) = E(Y|T)$  e  $X^{*T}\beta = E(Y|X^*)$ , pelo que estimá-la passará por estimar  $\beta$  e  $m$ .

Notemos ainda que, por definição,  $E(\varepsilon|X^*, T) = 0$ .

Multiplicando a equação (3.28) por  $\delta_i$  e tomando a esperança condicional em  $T_i = t$ , obtém-se

$$E(\delta_i Y_i | T_i = t) = E(\delta_i X_i^{*T} | T_i = t) \beta + E(\delta_i | T_i = t) m(t), \quad i = 1, \dots, n,$$

donde segue que

$$m(t) = m_2(t) - m_1(t)^T \beta, \quad (3.29)$$

com  $m_2(t) = \frac{E(\delta Y | T=t)}{E(\delta | T=t)}$  e  $m_1(t) = \frac{E(\delta X^{*T} | T=t)}{E(\delta | T=t)}$ .

Subtraindo a equação (3.29) multiplicada por  $\delta_i$  à equação (3.28) também multiplicada por  $\delta_i$ , obtém-se

$$\delta_i \{Y_i - m_2(T_i)\} = \delta_i \{X_i^* - m_1(T_i)\}^T \beta + \delta_i \varepsilon_i, \quad i = 1, \dots, n. \quad (3.30)$$

Somando os erros quadráticos, obtém-se

$$\sum_{i=1}^n \delta_i \varepsilon_i^2 = \sum_{i=1}^n \delta_i \{(Y_i - m_2(T_i)) - (X_i^* - m_1(T_i))^T \beta\}^2, \quad (3.31)$$

pelo que sentimo-nos tentados a estimar  $\beta$  pelo conhecido método dos mínimos quadrados. Isto é, minimizando a soma dos *erros* quadráticos relativamente ao parâmetro  $\beta$ . Contudo, a equação (3.31) depende ainda de outras duas quantidades desconhecidas,  $m_1(t)$  e  $m_2(t)$ . Assim, comecemos por estimar tais quantidades. Para tal, usemos o estimador *N-W* (introduzido na secção 2.2) associado ao conjunto de observações  $\{(X_i^*, T_i, Y_i), i = 1, \dots, n, \delta_i = 1\}$ . Deste modo, uma estimativa para  $m_1(t)$  será

$$\widehat{m_{1NW}}(t) = \frac{1}{n} \sum_{i=1}^n \delta_i W_{hi}(t) X_i^*$$

e uma para  $m_2(t)$  será

$$\widehat{m_{2NW}}(t) = \frac{1}{n} \sum_{i=1}^n \delta_i W_{hi}(t) Y_i,$$

onde o peso  $W_{hi}(t) = \frac{K_h(t-T_i)}{\frac{1}{n} \sum_{j=1}^n \delta_j K_h(t-T_j)}$ , com  $K_h(t-T_l) = \frac{1}{h} K\left(\frac{t-T_l}{h}\right)$ ,  $l = 1, \dots, n$ , e  $h$  é uma sucessão a decrescer para 0 enquanto  $n \rightarrow +\infty$ .

**Observação 3.2.** Notemos que o  $N-W$  associado ao conjunto  $\{(T_i, \delta_i), i = 1, \dots, n, \delta_i = 1\}$  atribui a  $E(\delta|T = t)$  uma estimativa igual 1.

Substituindo  $m_1(t)$  e  $m_2(t)$  em (3.31) pelas respectivas estimativas,  $\widehat{m}_{1NW}(t)$  e  $\widehat{m}_{2NW}(t)$ , uma estimativa para  $\beta$  será  $\widehat{\beta}_n$  que satisfaça

$$\min_{\beta} \sum_{i=1}^n \delta_i \varepsilon_i^2 \approx \min_{\beta} \psi(\beta), \quad (3.32)$$

com

$$\psi(\beta) = \sum_{i=1}^n \delta_i \left\{ (Y_i - \widehat{m}_{2NW}(T_i)) - (X_i^* - \widehat{m}_{1NW}(T_i))^T \beta \right\}^2,$$

baseado no conjunto  $\{(X_i^*, T_i, Y_i), i = 1, \dots, n, \delta_i = 1\}$ . Portanto, após derivar  $\psi$  (relativamente a  $\beta$ ) e igualar a 0, fácil computação permite-nos concluir que

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n \delta_i \{ (Y_i - \widehat{m}_{2NW}(T_i)) (X_i^* - \widehat{m}_{1NW}(T_i)) \}}{\sum_{i=1}^n \delta_i \{ (X_i^* - \widehat{m}_{1NW}(T_i)) (X_i^* - \widehat{m}_{1NW}(T_i))^T \}}. \quad (3.33)$$

Substituindo  $m_1(t)$ ,  $m_2(t)$  e  $\beta$  em (3.29) pelas respectivas estimativas, obtém-se uma estimativa para  $m(t)$ . Estimativa esta que é

$$\widehat{m}_n(t) = \widehat{m}_{2NW}(t) - \widehat{m}_{1NW}(t)^T \widehat{\beta}_n. \quad (3.34)$$

Desta forma, conclui-se também que uma estimativa para a *função de regressão* em  $(X^*, T) = (x, t)$  será

$$x^{*T} \widehat{\beta}_n + \widehat{m}_n(t).$$

### 3.3.2. Uma classe geral de estimadores

Após a estimação da *função de regressão* no modelo (3.28), (*MRPL*), estamos finalmente, em condições de estimar  $\theta$  quando há observações de  $Y$  *aleatoriamente em falta* no mesmo modelo.

Começemos por introduzir uma classe geral de estimadores

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{P_n^*(X_i^*, T_i)} Y_i + \left( 1 - \frac{\delta_i}{P_n^*(X_i^*, T_i)} \right) \left( X_i^{*T} \widehat{\beta}_n + \widehat{m}_n(T_i) \right) \right\}, \quad (3.35)$$

onde  $P_n^*(x, t)$  é uma sequência de quantidades com probabilidade limite  $P^*(x, t)$ .

Neste trabalho estamos particularmente interessados em alguns casos.

O primeiro destes,

$$\widehat{\theta}_I = \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i Y_i + (1 - \delta_i) \left( X_i^{*T} \widehat{\beta}_n + \widehat{m}_n(T_i) \right) \right\}, \quad (3.36)$$

que verifica-se quando  $P_n^*(x, t) = 1$ , simplesmente imputa uma estimativa  $\widehat{Y}_i = \left( X_i^{*T} \widehat{\beta}_n + \widehat{m}_n(T_i) \right)$  a todos os  $Y_i$ 's em falta e estima  $\theta$  naturalmente como se as  $n$

observações de  $Y$  estivessem completas. Este estimador assemelha-se ao estimador de Cheng introduzido na subsecção anterior, residindo a única diferença no facto de usarem diferentes modelos de regressão para as imputações das observações em falta. Designemo-lo por estimador da *imputação por regressão*.

O segundo caso,

$$\hat{\theta}_M = \frac{1}{n} \sum_{i=1}^n \left( X_i^{*T} \hat{\beta}_n + \widehat{m}_n(T_i) \right), \quad (3.37)$$

que verifica-se quando  $P_n^*(x, t) = \infty$ , apenas faz a média sobre as estimativas da *função de regressão* (nos pontos  $(X_1^*, T_1), \dots, (X_n^*, T_n)$ ), razão pela qual este estimador será designado por estimador da *média marginal*.

**Definição 3.1.** Definamos a probabilidade de  $Y$  ser observada dado  $T = t$ ,

$$P(\delta = 1|T = t) = E(\delta|T = t),$$

por *índice marginal de observação (i.m.o.)* e denotemo-la por  $P_1(t)$ .

O estimador  $N$ - $W$  permite-nos imputar a  $P_1(t)$  a estimativa (consistente, segundo a proposição 2.2)

$$\widehat{P}_1(t) = \frac{1}{n} \sum_{i=1}^n W_{hi}(t) \delta_i,$$

onde o *peso*  $W_{hi}(t) = \frac{K_h(t-T_i)}{\frac{1}{n} \sum_{j=1}^n K_h(t-T_j)}$ , com  $K_h(t - T_i) = \frac{1}{h} K\left(\frac{t-T_i}{h}\right)$ .

Quando  $P_n^*(x, t) = \widehat{P}_1(t)$  obtemos o estimador pesado com *i.m.o.*

$$\widehat{\theta}_{P_1} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{\widehat{P}_1(T_i)} Y_i + \left( 1 - \frac{\delta_i}{\widehat{P}_1(T_i)} \right) \left( X_i^{*T} \hat{\beta}_n + \widehat{m}_n(T_i) \right) \right\}. \quad (3.38)$$

Do ponto de vista estatístico e computacional, os estimadores  $\widehat{\theta}_I$ ,  $\widehat{\theta}_M$  e  $\widehat{\theta}_{P_1}$  não são difíceis de usar, uma vez que estão explicitamente definidos e dependem exclusivamente de simples cálculos. Relativamente a  $\widehat{\theta}_M$  e  $\widehat{\theta}_{P_1}$ ,  $\widehat{\theta}_I$  possui ainda a seguinte vantagem: caso a amostra dos  $Y_i$ 's disponíveis for aumentada com novas observações, estas podem ser directamente introduzidas na média dos mesmos, isto é, se considerarmos

$$\begin{aligned} \widehat{\theta}_I &= \frac{1}{n} \sum_{i=1}^n \delta_i Y_i + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left( X_i^{*T} \hat{\beta}_n + \widehat{m}_n(T_i) \right) \\ &= \widehat{\theta}_{I_1} + \widehat{\theta}_{I_2}, \end{aligned} \quad (3.39)$$

tais observações podem ser directamente introduzidas no cálculo de  $\widehat{\theta}_{I_1} = \frac{1}{n} \sum_{i=1}^n \delta_i Y_i$ .



**Definição 3.2.** Definamos a probabilidade de  $Y$  ser observada dado  $X^* = x$  e  $T = t$ ,

$$P(\delta = 1|X^* = x, T = t) = E(\delta|X^* = x, T = t),$$

por *índice (geral) de observação (i.o.)* e denotemo-la por  $P(x, t)$ .

O  $N-W$  permite-nos imputar a  $P(x, t)$  a estimativa (consistente, segundo a posição 2.2)

$$\hat{P}(x, t) = \frac{1}{n} \sum_{i=1}^n \left( \frac{\mathcal{K}\left(\frac{x-X_i^*}{b}, \frac{t-T_i}{b}\right)}{\frac{1}{n} \sum_{j=1}^n \mathcal{K}\left(\frac{x-X_j^*}{b}, \frac{t-T_j}{b}\right)} \right) \delta_i,$$

onde  $\mathcal{K}$  é a função *kernel* operando sobre argumento de dimensão  $d > 1$  e  $b = b(n)$  é uma seqüência a decrescer para 0 enquanto  $n \rightarrow +\infty$ . Assim, a juntar-se à classe de estimadores  $\hat{\theta}$ , tem-se ainda o estimador pesado com *i.o.*

$$\hat{\theta}_P = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{\hat{P}(X_i^*, T_i)} Y_i + \left( 1 - \frac{\delta_i}{\hat{P}(X_i^*, T_i)} \right) \left( X_i^{*T} \hat{\beta}_n + \hat{m}_n(T_i) \right) \right\}, \quad (3.40)$$

quando  $P_n^*(x, t) = \hat{P}(x, t)$ . Contudo, a *maldição da dimensionalidade* pode restringir o uso deste estimador.

Sejam  $g(x, t) = x^T \beta + m(t)$ ,  $\|\cdot\|$  a norma euclidiana,

$$\sigma^2(x, t) = Var(\varepsilon|X^* = x, T = t) = E \left[ (Y - X^{*T} \beta - m(T))^2 \middle| X^* = x, Y = t \right],$$

$P_1(t) = P(\delta = 1|T = t)$  e  $P(x, t) = P(\delta = 1|X^* = x, T = t)$ . Definamos  $u(x, t) = x - m_1(t)$ ,  $\Sigma = E [P(X^*, T)u(X^*, T)u(X^*, T)^T]$  e denotemos por  $m_{1r}$  à  $r$ -ésima componente de  $m_1$ .

O teorema que seguidamente enunciaremos garante a *normalidade assintótica* de  $\hat{\theta}$  para  $P_n^*(x, t) \in \{1, \infty, \hat{P}_1(t), \hat{P}(x, t)\}$ . Antes de enunciá-lo, consideremos as seguintes hipóteses:

$$(H_{13}) \sup_t E \left[ \|X^*\|^2 \middle| T = t \right] < \infty;$$

( $H_{14}$ )  $T$  admite densidade, diga-se  $f_T$ , e esta satisfaz

$$0 < \inf_{t \in [0, 1]} f_T(t) \leq \sup_{t \in [0, 1]} f_T(t) < +\infty,$$

isto é, a densidade de  $T$  é uma função limitada em  $[0, 1]$ ;

$$(H_{15}) \sup_t E [Y^2|X^* = x, T = t] < \infty;$$

( $H_{16}$ )  $m$ ,  $m_{1r}$  e  $m_2$  satisfazem a *condição de Lipschitz* de ordem 1;

(H<sub>17</sub>)  $\inf_{x,t} P(x, t) > 0$  e  $P_1(t)$  admite derivadas parciais de primeira e segunda ordem limitadas;

(H<sub>18</sub>)  $\Sigma$  é uma matrix *definida positiva*;

(H<sub>19</sub>)  $K$  é uma função *kernel* de segunda ordem a satisfazer  $M_1 I_{[|s| \leq \rho]} \leq K(s) \leq M_2 I_{[|s| \leq \rho]}$ , para algumas constantes  $M_1 > 0$ ,  $M_2 > 0$  e  $\rho > 0$  (o que implica que  $K$  é limitada, de suporte limitado e estritamente positiva);

(H<sub>20</sub>) A função *kernel* operando sobre argumento de dimensão  $d > 1$ ,  $\mathcal{K}$ , tem suporte e variação limitados e é de ordem  $k > d + 1$ , isto é, tem momentos de ordem 1 até  $d + 1$  nulos e de ordem  $k$  não nulo e finito;

(H<sub>21</sub>)  $nh \rightarrow +\infty$  e  $nh^2 \rightarrow 0$ ;

(H<sub>22</sub>)  $\frac{nb^{2(d+1)}}{\log(n)} \rightarrow +\infty$  e  $nb^{2k} \rightarrow 0$ .

**Teorema 3.3.** *Assumindo as hipóteses (H<sub>15</sub>) a (H<sub>22</sub>) acima listadas e o MRPL, tem-se*

$$(i) \sqrt{n} \left( \hat{\theta}_I - \theta \right) \xrightarrow{L} \mathcal{N}(0, V);$$

$$(ii) \sqrt{n} \left( \hat{\theta}_M - \theta \right) \xrightarrow{L} \mathcal{N}(0, V);$$

$$(iii) \sqrt{n} \left( \hat{\theta}_{P_1} - \theta \right) \xrightarrow{L} \mathcal{N}(0, V);$$

$$(iv) \sqrt{n} \left( \hat{\theta}_P - \theta \right) \xrightarrow{L} \mathcal{N}(0, V');$$

Onde  $V = \mathbb{E} \left[ [\pi_0(X^*, T) + \pi_1(X^*, T)]^2 P(X^*, T) \sigma^2(X^*, T) \right] + \text{Var}(g(X^*, T))$ , com  $\pi_0(x, t) = \frac{1}{P_1(t)}$  e  $\pi_1(x, t) = \mathbb{E} \left[ u(X^*, T)^T \right] \frac{1}{\Sigma} u(x, t)$ ,  $V' = \mathbb{E} \left[ \frac{\sigma^2(X^*, T)}{P(X^*, T)} \right] + \text{Var}(g(X^*, T))$  e  $\mathcal{N}(0, \Delta)$  é a lei normal de média nula e variância  $\Delta$ .

*Demonstração.* Esboçaremos a prova para cada um dos casos (i), (ii), (iii) e (iv).

Para (i), tem-se

$$\begin{aligned} \hat{\theta}_I &= \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i Y_i + (1 - \delta_i) (X_i^{*T} \beta + m(T_i)) \right\} \\ &\quad + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) X_i^{*T} \left( \hat{\beta}_n - \beta \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) (\hat{m}_n(T_i) - m(T_i)) \\ &= A_{n1} + A_{n2} + A_{n3}. \end{aligned} \tag{3.41}$$

Vamos à subsecção 3.3.1 buscar a expressão de  $\widehat{\beta}_n$  para concluir que

$$\begin{aligned}
 \widehat{\beta}_n - \beta &= \frac{\sum_{i=1}^n \delta_i [(Y_i - \widehat{m}_{2NW}(T_i)) [X_i^* - \widehat{m}_{1NW}(T_i)]}{\sum_{i=1}^n \delta_i [X_i^* - \widehat{m}_{1NW}(T_i)] [X_i^* - \widehat{m}_{1NW}(T_i)]^T} - \beta \\
 &= \frac{\sum_{i=1}^n \delta_i [X_i^* - \widehat{m}_{1NW}(T_i)] [Y_i - X_i^* \beta - \widehat{m}_n(T_i)]}{\sum_{i=1}^n \delta_i [X_i^* - \widehat{m}_{1NW}(T_i)] [X_i^* - \widehat{m}_{1NW}(T_i)]^T} \\
 &= \frac{1}{n\Sigma} \sum_{i=1}^n \delta_i [X_i^* - m_1(T_i)] \varepsilon_i + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.42}
 \end{aligned}$$

Substituindo  $Y_i$  por  $X_i^{*T}\beta + m(T_i) + \varepsilon_i$ , usando as expressões de  $\widehat{m}_{2NW}(t)$  e  $\widehat{m}_{1NW}(t)$  (subsecção 3.3.1) e as igualdades  $\lim_{h \rightarrow 0} \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) m(T_j) = m(T_i)$ ,  $\lim_{h \rightarrow 0} \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) \varepsilon_j = \varepsilon_i$ ,  $i = 1, \dots, n$ , e  $E\left(\frac{\delta}{P_1(T)}\right) = 1$ , obtemos

$$\begin{aligned}
 A_{n3} &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) (\widehat{m}_n(T_i) - m(T_i)) \\
 &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left[ \widehat{m}_{2NW}(T_i) - \widehat{m}_{1NW}(T_i)^T \widehat{\beta}_n - m(T_i) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left[ \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) Y_j - \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) X_j^{*T} \widehat{\beta}_n - m(T_i) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left[ \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) \varepsilon_j - \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) X_j^{*T} (\widehat{\beta}_n - \beta) \right] \\
 &\quad + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \left[ \varepsilon_i - \widehat{m}_{1NW}(T_i)^T (\widehat{\beta}_n - \beta) \right] + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i - \frac{1}{n} \sum_{i=1}^n \varepsilon_i \delta_i - \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) m_1(T_i)^T (\widehat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{1}{n} \sum_{i=1}^n \delta_i \varepsilon_i \frac{(1 - P_1(T_i))}{P_1(T_i)} - \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) m_1(T_i)^T (\widehat{\beta}_n - \beta) \\
 &\quad + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.43}
 \end{aligned}$$

Substituindo  $Y_i$  por  $X_i^{*T}\beta + m(T_i) + \varepsilon_i$  em

$$A_{n1} = \frac{1}{n} \sum_{i=1}^n \{ \delta_i Y_i + (1 - \delta_i) (X_i^{*T} \beta + m(T_i)) \},$$

as igualdades (3.41), (3.42) e (3.43) permitem obter

$$\begin{aligned}
 \widehat{\theta}_I - \theta &= \frac{1}{n} \sum_{i=1}^n \delta_i \varepsilon_i + \frac{1}{n} \sum_{i=1}^n (X_i^{*T} \beta + m(T_i) - \theta) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) X_i^{*T} (\widehat{\beta}_n - \beta) + \frac{1}{n} \sum_{i=1}^n \delta_i \varepsilon_i \left( \frac{1}{P_1(T_i)} - 1 \right) \\
 &\quad - \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) m_1(T_i)^T (\widehat{\beta}_n - \beta) + o_p \left( \frac{1}{\sqrt{n}} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{P_1(T_i)} + E[u(X^*, T)^T] \frac{1}{\Sigma} \delta_i (X_i^* - m_1(T_i)) \right\} \varepsilon_i \\
 &\quad + \frac{1}{n} \sum_{i=1}^n (X_i^{*T} \beta + m(T_i) - \theta) + o_p \left( \frac{1}{\sqrt{n}} \right), \tag{3.44}
 \end{aligned}$$

Para não sobrecarregar as notações, em (3.45) e (3.46), a  $u(X^*, T)$  denotaremos por  $u$ . Computação directa permite-nos concluir que  $E \left[ \sqrt{n} (\widehat{\theta}_I - \theta) \right] = 0$ . De facto,

$$\begin{aligned}
 E(\widehat{\theta}_I - \theta) &= E \left[ \frac{\delta \varepsilon}{P_1(T)} \right] + E \left[ E(u)^T \frac{1}{\Sigma} u \delta \varepsilon \right] + E[E(Y|X^*, T) - \theta] \\
 &= E \left[ \frac{1}{P_1(T)} E(\delta|X^*, T) E(\varepsilon|X^*, T) \right] \\
 &\quad + E \left[ E(u)^T \frac{1}{\Sigma} u E(\delta|X^*, T) E(\varepsilon|X^*, T) \right] \\
 &= 0. \tag{3.45}
 \end{aligned}$$

Para a variância assintótica,  $Var \left( \sqrt{n} (\widehat{\theta}_I - \theta) \right)$ , notemos que os termos em (3.44) são não correlacionados. Deste modo, tem-se

$$\begin{aligned}
 nVar \left( (\widehat{\theta}_I - \theta) \right) &= Var \left( \frac{\delta \varepsilon}{P_1(T)} + E(u)^T \frac{1}{\Sigma} u \delta \varepsilon \right) + Var(g(X^*, T)) \\
 &= E \left[ \frac{1}{P_1^2(T)} E(\delta|X^*, T) Var(\varepsilon|X^*, T) \right] \\
 &\quad + 2E \left[ \frac{1}{P_1(T)} E(u)^T \frac{1}{\Sigma} u E(\delta|X^*, T) Var(\varepsilon|X^*, T) \right] \\
 &\quad + E \left[ E^2(u)^T \frac{1}{\Sigma^2} u^2 E(\delta|X^*, T) Var(\varepsilon|X^*, T) \right] \\
 &\quad + Var(g(X^*, T)), \tag{3.46}
 \end{aligned}$$

pelo que  $nVar \left( (\widehat{\theta}_I - \theta) \right) = V$  conforme o enunciado do teorema. Nestas condições, o teorema do limite central assegura a normalidade assintótica de  $\widehat{\theta}_I$ .

Para (ii), tem-se

$$\begin{aligned}
 \widehat{\theta}_M - \theta &= \frac{1}{n} \sum_{i=1}^n (X_i^{*T} \beta - m(T_i) - \theta) + \frac{1}{n} \sum_{i=1}^n X_i^{*T} (\widehat{\beta}_n - \beta) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n (\widehat{m}_n(T_i) - m(T_i)) \\
 &= B_{n1} + B_{n2} + B_{n3}. \tag{3.47}
 \end{aligned}$$

Notemos que

$$\begin{aligned} B_{n2} &= \frac{1}{n} \sum_{i=1}^n X_i^{*T} (\hat{\beta}_n - \beta) \\ &= E(X^*)^T (\hat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (3.48)$$

Notemos ainda que, procedendo-se de forma idêntica ao caso (3.43), obtém-se

$$\begin{aligned} B_{n3} &= \frac{1}{n} \sum_{i=1}^n (\widehat{m}_n(T_i) - m(T_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) \varepsilon_j - \frac{1}{n} \sum_{j=1}^n \delta_j W_{hi}(T_i) X_j^{*T} (\hat{\beta}_n - \beta) \right] \\ &\quad + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i - \frac{1}{n} \sum_{i=1}^n m_{1NW}(T_i)^T (\hat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{\delta_i}{P_1(T_i)} - E(m_1(T))^T (\hat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (3.49)$$

Assim, pelas equações (3.44), (3.47), (3.48) e (3.49) concluímos que

$$\hat{\theta}_M = \hat{\theta}_I + o_p\left(\frac{1}{\sqrt{n}}\right) \quad (3.50)$$

e conseqüentemente a *normalidade assintótica* de  $\hat{\theta}_M$ , conforme o teorema.

Para (iii), tem-se

$$\begin{aligned} \hat{\theta}_{P_1} - \theta &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varepsilon_i}{\widehat{P}_1(T_i)} - \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\widehat{P}_1(T_i)} X_i^{*T} (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{n} \sum_{i=1}^n (X_i^{*T} \hat{\beta}_n + m_n(T_i)) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\widehat{P}_1(T_i)} (m(T_i) - m_n(T_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varepsilon_i}{P_1(T_i)} + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varepsilon_i (\widehat{P}_1(T_i) - P_1(T_i))}{P_1^2(T_i)} \\ &\quad + \frac{1}{n} \sum_{i=1}^n (X_i^{*T} \beta + m(T_i) - \theta) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) X_i^{*T} (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) (m_n(T_i) - m(T_i)) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= C_{n1} + C_{n2} + C_{n3} + C_{n4} + C_{n5} + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \quad (3.51)$$

somando e subtraindo os termos  $\frac{1}{n} \sum_{i=1}^n m(T_i)$ ,  $\frac{1}{n} \sum_{i=1}^n X_i^{*T} \beta$  e  $\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varepsilon_i}{P_1(T_i)}$ .

Para  $C_{n5}$ , notemos que  $E\left(1 - \frac{\delta}{P_1(T)} \middle| T_i\right) = 0$ . Deste modo, pela equação (3.43) e pelo facto de  $\widehat{\beta}_n - \beta = o_p\left(\frac{1}{\sqrt{n}}\right)$ , tem-se

$$\begin{aligned}
 C_{n5} &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) (m_n(T_i) - m(T_i)) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) \frac{1}{P_1(T_i)} \left[\delta_i \varepsilon_i - \delta_i m_1(T_i)^T (\widehat{\beta}_n - \beta)\right] \\
 &\quad + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) \frac{1}{P_1(T_i) f_T(T_i)} \sum_{j=1}^n K_h(T_i - T_j) \delta_j \varepsilon_j \\
 &\quad - \frac{1}{n^2} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) \frac{1}{P_1(T_i) f_T(T_i)} \sum_{j=1}^n K_h(T_i - T_j) \delta_j m_1(T_j)^T \\
 &\quad \times (\widehat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{K(0)}{hn^2} \sum_{j=1}^n \frac{\delta_j \varepsilon_j}{P_1(T_j) f_T(T_j)} \left(1 - \frac{\delta_j}{P_1(T_j)}\right) \\
 &\quad + \frac{K(0)}{hn^2} \sum_{j=1}^n \left(1 - \frac{\delta_j}{P_1(T_j)}\right) \frac{\delta_j m_1(T_j)^T}{P_1(T_j) f_T(T_j)} (\widehat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= o_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.52}
 \end{aligned}$$

Para  $C_{n4}$ , tem-se

$$\begin{aligned}
 C_{n4} &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P_1(T_i)}\right) X_i^{*T} (\widehat{\beta}_n - \beta) \\
 &= E(X^{*T} - m_1(T))^T (\widehat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= E(u(X^*, T)^T) (\widehat{\beta}_n - \beta) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.53}
 \end{aligned}$$

Para  $C_{n2}$ , tem-se

$$\begin{aligned}
 C_{n2} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varepsilon_i}{P_1^2(T_i)} (\widehat{P}_1(T_i) - P_1(T_i)) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \frac{\delta_i \varepsilon_i}{P_1^2(T_i) f_T(T_i)} \left( \sum_{j=1}^n K_h(T_i - T_j) \delta_j - \sum_{j=1}^n K_h(T_i - T_j) P_1(T_j) \right) \\
 &\quad + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= \frac{1}{hn^2} \sum_{j=1}^n (\delta_j - P_1(T_j)) \sum_{i=1}^n K\left(\frac{T_i - T_j}{h}\right) \frac{\delta_i \varepsilon_i}{P_1^2(T_i) f_T(T_i)} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &= o_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.54}
 \end{aligned}$$

Finalmente, por (3.51), (3.52), (3.53) e (3.54) concluímos que

$$\widehat{\theta}_{P_1} = \widehat{\theta}_I + o_p\left(\frac{1}{\sqrt{n}}\right) \quad (3.55)$$

e consequentemente a *normalidade assintótica* de  $\widehat{\theta}_{P_1}$  conforme o teorema.

Para concluir a demonstração, provemos o caso (iv). Para simplificar as notações, seja  $Z_i = (X_i^*, Y_i)$ . Ora,

$$\begin{aligned} \widehat{\theta}_P &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i Y_i}{\widehat{P}(Z_i)} + \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\widehat{P}(Z_i)}\right) \left(X_i^{*T} \widehat{\beta}_n + \widehat{m}_n(T_i)\right) \\ &= D_{n1} + D_{n2}. \end{aligned} \quad (3.56)$$

Provemos que

$$D_{n2} = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\widehat{P}(Z_i)}\right) \left(X_i^{*T} \widehat{\beta}_n + \widehat{m}_n(T_i)\right) \longrightarrow 0. \quad (3.57)$$

De facto, pelas condições  $(H_{14})$ ,  $(H_{19})$ ,  $(H_{20})$ ,  $(H_{21})$  e  $(H_{22})$ , obtém-se

$$\begin{aligned} D_{n2} &= \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P(Z_i)}\right) \left(X_i^{*T} \beta + m(T_i)\right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \left(\widehat{P}(Z_i) - P(Z_i)\right)}{P^2(Z_i)} \left(X_i^{*T} \beta + m(T_i)\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \quad (3.58)$$

onde, usando a mesma técnica que em (3.54), conclui-se que

$$\begin{aligned} D_{n2}^* &= \frac{1}{n} \sum_{j=1}^n (\delta_j - P(Z_j)) \frac{1}{nb^{d+1}} \sum_{i=1}^n \frac{\delta_i}{P^2(Z_i)g(Z_i)} \mathcal{K}\left(\frac{Z_i - Z_j}{b}\right) \\ &\quad \times \left(X_i^{*T} \beta + m(T_i)\right) \\ &= \frac{1}{n} \sum_{j=1}^n \frac{\delta_j - P(Z_j)}{P(Z_j)} \left(X_j^{*T} \beta + m(T_j)\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \quad (3.59)$$

onde  $D_{n2}^*$  é o segundo termo do lado direito da equação (3.58),  $g$  é a função densidade de  $Z$  e  $\widehat{g}_n(z) = \frac{1}{nb^{d+1}} \sum_{i=1}^n \mathcal{K}\left(\frac{z - Z_i}{b}\right)$ .

Pelas equações (3.58) e (3.59) concluímos o resultado (3.57) e, como consequência,

$$\begin{aligned} \widehat{\theta}_P &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i Y_i}{\widehat{P}(Z_i)} + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \widehat{\theta}_{HIR} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad (3.60)$$

$\widehat{\theta}_{HIR} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i Y_i}{\widehat{P}(Z_i)}$  foi introduzido por Hirano, Imbens, and Ridder [?] para estimar o efeito médio de um determinado tipo de tratamento nos indivíduos duma população de tamanho  $n$ . Hirano, Imbens, and Ridder [?] provaram que  $\sqrt{n} \left(\widehat{\theta}_{HIR} - \theta\right)$  converge para *lei normal* de média nula e variância  $V' = E\left[\frac{\sigma^2(Z)}{P(Z)}\right] + Var(g(Z))$ . Concluímos assim a demonstração do teorema que também pode ser encontrada em Wang, Linton, and Härdle [?].  $\square$

A demonstração do teorema anterior permite-nos concluir que os estimadores  $\widehat{\theta}_I$ ,  $\widehat{\theta}_M$  e  $\widehat{\theta}_{P_1}$  são assintoticamente equivalentes, isto é, diferem numa quantidade negligenciável quando  $n \rightarrow +\infty$ . Adoptemos a designação comum  $\widehat{\theta}_*$  para estes estimadores ( $\widehat{\theta}_I$ ,  $\widehat{\theta}_M$  e  $\widehat{\theta}_{P_1}$ ).

Para definir estimadores consistentes para as variâncias assintóticas  $V$  e  $V'$ , vamos usar o estimador Jackknife para a variância. Uma outra alternativa será estimar as quantidades desconhecidas  $\sigma^2(x, t)$ ,  $P_1(t)$ ,  $P(x, t)$  e  $m_1(t)$  usando o estimador  $N-W$  e definir estimadores consistentes para  $V$  e  $V'$ , pelo método “plug-in”. Pois, dado o ponto  $(x, t)$ , as únicas quantidades desconhecidas nas expressões de  $V$  e  $V'$  são as que acabamos de referir. Contudo, quando a dimensão do espaço onde  $X = (X^*, T)$  toma valores for grande, a *maldição da dimensionalidade* pode levar-nos a construção de estimativas com pouca ou nenhuma fiabilidade para tais quantidades, sobretudo nos casos de  $\sigma^2(x, t)$  e  $P(x, t)$ .

**Definição 3.3.** Sejam  $\widehat{\theta}_*^{-i}$  e  $\widehat{\theta}_P^{-i}$ , respectivamente, os estimadores  $\widehat{\theta}_*$  e  $\widehat{\theta}_P$  baseados no conjunto de observações  $\{(Y_j, \delta_j, X_j^*, T_j), j = 1, \dots, i-1, i+1, \dots, n\}$ . Sejam ainda os pseudo-estimativas de Jackknife  $J_{ni} = n\widehat{\theta}_* - (n-1)\widehat{\theta}_*^{-i}$ ,  $i = 1, \dots, n$ , e  $J'_{ni} = n\widehat{\theta}_P - (n-1)\widehat{\theta}_P^{-i}$ ,  $i = 1, \dots, n$ . Os estimadores de Jackknife para  $V$  e  $V'$  são respectivamente  $\widehat{V}_{nJ} = \frac{1}{n} \sum_{i=1}^n (J_{ni} - \bar{J}_n)^2$  e  $\widehat{V}'_{nJ} = \frac{1}{n} \sum_{i=1}^n (J'_{ni} - \bar{J}'_n)^2$ , onde  $\bar{J}_n = \frac{1}{n} \sum_{i=1}^n J_{ni}$  e  $\bar{J}'_n = \frac{1}{n} \sum_{i=1}^n J'_{ni}$ .

Os estimadores  $\widehat{V}_{nJ}$  e  $\widehat{V}'_{nJ}$ , pelas suas consistências, conforme o teorema 3.2 de Wang, Linton and Härdle [?], constituem também ótimas ferramentas na inferência sobre o intervalo de confiança e teste de hipóteses para  $\theta$ . De facto, pelo teorema 3.3 deste trabalho, para  $\widehat{\theta}_*$ , por exemplo,  $\frac{\widehat{\theta}_* - \theta}{\sqrt{\frac{V}{n}}}$  converge para a *lei normal standard* e, sendo  $\widehat{V}_{nJ}$  uma estimativa consistente de  $V$ , conforme o teorema 3.2 de Wang, Linton and Härdle [?], o intervalo  $\left[ \widehat{\theta}_* - \sqrt{\frac{\widehat{V}_{nJ}}{n}} s_{\frac{\alpha}{2}}, \widehat{\theta}_* + \sqrt{\frac{\widehat{V}_{nJ}}{n}} s_{\frac{\alpha}{2}} \right]$  é um intervalo de confiança para  $\theta$  ao nível  $1-\alpha$ , onde  $s_{\frac{\alpha}{2}}$  é o quantil de ordem  $\frac{\alpha}{2}$  da *lei normal standard*. Relativamente a teste de hipóteses, pela razão enunciada no final da secção 3.2, não faremos nenhuma referência, pelo que, para mais detalhes, convidamos o leitor a consultar Lehmann [?].

### 3.4. Eficiência dos estimadores

Nas duas secções anteriores introduzimos alguns estimadores para  $\theta$  que, conforme podemos constatar, têm todos um bom desempenho assintótico. Neste sentido,



sentimo-nos tentados a comparar a eficiência dos mesmos.

Ora, na subsecção 3.3.2 provamos que os nossos estimadores  $\widehat{\theta}_I$ ,  $\widehat{\theta}_M$  e  $\widehat{\theta}_{P_1}$  têm todos a distribuição assintótica normal de média nula e variância

$$V = E \left[ \left( \frac{1}{P_1(T)} + E[u(X^*, T)^T] \frac{1}{\Sigma} u(X^*, T) \right)^2 P(X^*, T) \sigma^2(X^*, T) \right] + Var(g(X^*, T))$$

e denotamo-los por  $\widehat{\theta}_*$ . Consideremos a classe de estimadores

$$\widetilde{\theta} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{P_n^*(X_i^*, T_i)} Y_i + \left( 1 - \frac{\delta_i}{P_n^*(X_i^*, T_i)} \right) \widehat{m}_{NW}(X_i^*, T_i) \right\}, \quad (3.61)$$

onde  $\widehat{m}_{NW}(x, t)$  é o estimador *N-W* para a *função de regressão*  $E(Y|X^*, T)$  no ponto  $(x, t)$  e  $P_n^*(x, t)$  é a sequência com probabilidade limite  $P^*(x, t)$  introduzida em (3.35). Esta classe inclui entre outros o estimador de Chang [?],

$$\widehat{\theta}_{ch} = \frac{1}{n} \sum_{i=1}^n \{ \delta_i Y_i + (1 - \delta_i) \widehat{m}_{NW}(X_i^*, T_i) \},$$

quando  $P_n^*(x, t) = 1$ .

Pelos teoremas 3.1 e 3.3, sendo o teorema 3.1 nas conformidades com a nota 3.1, os estimadores  $\widehat{\theta}_{ch}$  e  $\widehat{\theta}_P$  têm ambos a distribuição assintótica normal de média nula e variância  $V_{cP}$ , onde  $V_{cP} = E \left[ \frac{\sigma^2(X^*, T)}{P(X^*, T)} \right] + Var(E(Y|X^*, T))$ . Denotemo-los por  $\widetilde{\theta}_*$ .

Em toda esta secção vamos assumir verdadeiro o *MRPL*, isto é,  $Y = \beta^T X^* + m(T) + \varepsilon$ , onde  $\beta^T X^* + m(T) = E(Y|X^*, T)$ . Obviamente, este modelo tem como um dos casos particulares o *MRNP*, que acontece quando  $X = T$ .

Antes de introduzir qualquer resultado que estabeleça a comparação entre os dois grupos de estimadores,  $\widehat{\theta}_*$  e  $\widetilde{\theta}_*$ , podemos dizer que, intuitivamente, relativamente a  $\widehat{\theta}_*$ ,  $\widetilde{\theta}_*$  possui a desvantagem de requerer operações sobre argumentos que pertencem a espaços de maiores dimensões para computar a regressão de  $Y$  sobre  $(X^*, T)$  e, portanto, corre o risco de ter um menor desempenho. Pois,  $\widetilde{\theta}_*$  depende das estimativas de *N-W*,  $\widehat{m}_{NW}(x, t)$  e  $\widehat{P}(x, t)$ , dentre outras, que por causa da *maldição da dimensionalidade* (ver secção 1.2), poderão não ter um bom desempenho quando o argumento  $(x, t)$  pertence a um espaço de dimensão grande.

Os estimadores  $\widehat{\theta}_*$ , conforme podemos constatar, dependem de estimativas de *N-W* operando sobre argumentos de dimensão no máximo  $d^* (\leq d)$ , ou melhor, estes dependem de estimativas de *N-W* para  $P_1(t)$  e  $m(t)$ , que podem ser eficientemente construídos dado que o estimador *N-W* tem um melhor desempenho quando aplicado a argumentos provenientes de espaços de dimensões pequenas.

O teorema seguinte compara a eficiência dos nossos estimadores  $\widehat{\theta}_*$  com os estimadores  $\widetilde{\theta}_*$  quando a variância condicional de  $\varepsilon$  dado  $(X^* = x, T = t)$ ,  $\sigma^2(x, t)$ , é constante.

**Observação 3.3.** Computação directa permiti-nos concluir que

$$\begin{aligned} \text{Var}(\varepsilon|X^* = x, T = t) &= \text{E} \left[ (Y - X^{*T}\beta - m(T))^2 \mid X^* = x, Y = t \right] \\ &= \text{Var}(Y|X^* = x, T = t). \end{aligned} \quad (3.62)$$

**Teorema 3.4.** Se  $\sigma^2(x, t) = \nu^2$ , onde  $\nu^2$  é uma constante, então

$$V \leq V_{cP}. \quad (3.63)$$

A igualdade verifica-se quando

$$\left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \varepsilon = a\delta(X^* - m_1(T))\varepsilon + b,$$

para algumas constantes  $a$  e  $b$ .

*Demonstração.* Nas condições do teorema, tem-se

$$V = \nu^2 \text{E} \left[ \frac{1}{P_1(T)} \right] + \nu^2 \text{E} [u(X^*, T)^T] \frac{1}{\Sigma} \text{E} [u(X^*, T)] + \text{Var}(g(X^*, T)) \quad (3.64)$$

e

$$V_{cP} = \nu^2 \text{E} \left[ \frac{1}{P(X^*, T)} \right] + \text{Var}(g(X^*, T)). \quad (3.65)$$

Notemos que

$$\begin{aligned} \text{E} \left[ \frac{\delta}{P_1(T)} (X^* - m_1(T)) \right] &= \text{E} \left[ \frac{\text{E}(\delta X^*|T)}{P_1(T)} - \text{E}(\delta X^*|T) \frac{\text{E}(\delta|T)}{P_1^2(T)} \right] \\ &= 0, \end{aligned}$$

$$\begin{aligned} \text{E} \left[ \frac{\delta}{P(X^*, T)} (X^* - m_1(T)) \right] &= \text{E} \left[ (X^* - m_1(T)) \frac{\text{E}(\delta|X^*, T)}{P(X^*, T)} \right] \\ &= \text{E} [X^* - m_1(T)] \end{aligned}$$

e  $\text{E} [\delta(X^* - m_1(T))\varepsilon] = 0$ , pelo que

$$\begin{aligned} \nu^2 \text{E} [u(X^*, T)] &= \nu^2 \text{E} \left[ \left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \delta (X^* - m_1(T)) \right] \\ &= \text{Cov} \left[ \left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \varepsilon, \delta (X^* - m_1(T)) \varepsilon \right]. \end{aligned} \quad (3.66)$$

Além disso,

$$\frac{(\nu^2 \text{E} [u(X^*, T)^T]) (\nu^2 \text{E} [u(X^*, T)])}{\text{Var} \left( \left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \varepsilon \right) \text{Var} (\delta (X^* - m_1(T)) \varepsilon)} \leq 1 \quad (3.67)$$

e portanto

$$\begin{aligned} \frac{(\nu^2 \mathbb{E} [u(X^*, T)^T]) (\nu^2 \mathbb{E} [u(X^*, T)])}{\nu^2 \Sigma} &\leq \text{Var} \left( \left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \varepsilon \right) \\ &= \nu^2 \mathbb{E} \left[ \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right], \end{aligned} \quad (3.68)$$

uma vez que

$$\begin{aligned} \text{Var} (\delta (X^* - m_1(T)) \varepsilon) &= \nu^2 \mathbb{E} \left[ (X^* - m_1(T)) (X^* - m_1(T))^T \mathbb{E}[\delta | X^*, T] \right] \\ &= \nu^2 \Sigma. \end{aligned} \quad (3.69)$$

Reescrevendo de modo equivalente a inequação (3.68), obtêm-se

$$\nu^2 \mathbb{E} [u(X^*, T)^T] \frac{1}{\Sigma} \mathbb{E} [u(X^*, T)] + \nu^2 \mathbb{E} \left[ \frac{\delta}{P_1(T)} \right] \leq \nu^2 \mathbb{E} \left[ \frac{\delta}{P(X^*, T)} \right], \quad (3.70)$$

o que conclui a demonstração do teorema, isto é,  $V \leq V_{cP}$ . Claramente, a igualdade ( $V \leq V_{cP}$ ) verifica-se somente quando  $\left( \frac{\delta}{P(X^*, T)} - \frac{\delta}{P_1(T)} \right) \varepsilon = a\delta(X^* - m_1(T))\varepsilon + b$ . De facto, sob esta hipótese,

$$\text{Cov} [a\delta (X^* - m_1(T)) \varepsilon + b, \delta (X^* - m_1(T)) \varepsilon] = a\nu^2 \Sigma$$

e

$$\text{Var}(a\delta (X^* - m_1(T)) \varepsilon + b) = a^2 \nu^2 \Sigma,$$

pelo que reorganizando (3.67), verifica-se a igualdade  $V = V_{cP}$ .  $\square$

O teorema anterior confirma a nossa ideia intuitiva sobre uma maior eficiência dos estimadores  $\hat{\theta}_*$ , relativamente aos estimadores  $\tilde{\theta}_*$ .

### 3.5. Nota conclusiva

A estimação da média quando não se conhece a amostra na sua totalidade, vem confirmar a poderosa estrutura dos modelos de regressão bem como do estimador  $N-W$ . O modelo de regressão totalmente não paramétrico constitui uma óptima ferramenta na infêrencia sobre relações entre variáveis aleatórias. Contudo este modelo é relativamente limitado quando se confronta com a *maldição da dimensionalidade* (ver a secção 1.2).

Neste sentido, torna-se evidente que as estimativas imputadas às observações em falta pelo estimador de Cheng [?] (ver secção 3.2) não fornecem grande fiabilidade se a dimensão do espaço onde toma valores o vector aleatório  $X$  for relativamente grande, devido ao uso de estimador não paramétrico, que é neste caso o  $N-W$ . Claro

que, quando a dimensão do espaço for pequena, melhor ainda se for igual a 1, o estimador de Cheng [?] permitirá inferir fiáveis estimativas a tais observações. De um modo geral, podemos dizer que o estimador de Cheng [?] tem um bom desempenho quando a dimensão onde toma valores o vector das variáveis explicativas  $X$ , for pequena.

Por outro lado, tendo a noção de que algumas das margens em  $X = (X^*, T)$  influenciarão linearmente a variável (resposta)  $Y$ , o *MRPL* (ver secção 3.3) será uma excelente alternativa ao modelo completamente não paramétrico, na medida em que, modelando parametricamente estas margens e de forma completamente não paramétrica as restantes, os estimadores não paramétricos serão aplicados exclusivamente à estimação da parte não paramétrica ( $E(Y|T = t) = m(t)$ ), requerendo assim operações sobre argumentos de menores dimensões. Isto é, as observações às quais tais estimadores serão associados proverão de vectores aleatórios ( $T$ , no nosso caso) cuja dimensão do espaço onde tomam valores é sempre menor ou igual à dimensão do espaço onde toma valores o vector  $X$ . Quanto à parte paramétrica que figura no *MRPL*, como podemos constatar na secção 1.1 e na subsecção 3.3.1, esta é trivialmente tratada.

Para um estudo mais geral sobre os modelos de regressão que envolvam combinações de técnicas das abordagens paramétrica e não paramétrica, usualmente denominados por *modelos de regressão semiparamétricos*, recomendamos o Härdle, Müller, Sperlich and Werwatz [?].