

# Avaliação de um algoritmo adaptativo para previsão de taxas de perda de pacotes em redes de computadores



UFRJ

Dissertação submetida para a obtenção do título de

**Mestre em Ciências em**

**Engenharia de Sistemas e Computação**

ao Programa de Pós-Graduação de Engenharia de Sistemas e Computação

da COPPE/UFRJ

por

**Hugo Hidequi Costa Sato**

Dezembro 2007

AVALIAÇÃO DE UM ALGORITMO ADAPTATIVO PARA PREVISÃO DE  
TAXAS DE PERDA DE PACOTES EM REDES DE COMPUTADORES

Hugo Hidequi Costa Sato

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO  
DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE  
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

---

Prof. Edmundo Albuquerque de Souza e Silva, Ph.D.

---

Prof.<sup>a</sup> Rosa Maria Meri Leão, Dr.

---

Prof. Daniel Ratton Figueiredo, Ph.D.

---

Prof. Artur Ziviani, Dr.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2007

SATO, HUGO HIDEQUI COSTA

Avaliação de um algoritmo adaptativo para  
previsão de taxas de perda de pacotes em re-  
des de computadores [Rio de Janeiro] 2007

xii, 74 p. 29,7 cm (COPPE/UFRJ, M.Sc.,  
Engenharia de Sistemas e Computação, 2007)

Dissertação - Universidade Federal do Rio  
de Janeiro, COPPE

1. Modelos de Markov ocultos
  2. Previsão de perdas de pacotes
  3. Análise e previsão de séries temporais
- I. COPPE/UFRJ      II. Título (Série)

# Agradecimentos

Escrevo esta seção antes das demais – acredite! – por um desengano de consciência. Preciso registrar meus agradecimentos e reconhecimento àqueles que, ao longo destes anos, me ajudaram a chegar até aqui.

Dedico este trabalho aos meus pais e irmãos, Silvia, Diogo, Thiago e Caio . Sem menos merecimento, também dedico à minha companheira nesta jornada: Julia Paula. A eles, reconheço uma eterna gratidão pela paciência com meu “temperamento” durante estes anos.

Pelo companheirismo, aos amigos Fabianne e Guto, GD, Fernando, Watanabe, Ana Paula, Carol e João, Bene, Boechat, Bernardo, Bruno, Sadoc, Fabrício, Ligia e Paulo, Tulio; meus sinceros agradecimentos. Um agradecimento especial à Carolina Maria de Vilar Moreira Vieira, mais alegres são nossos dias no laboratório com a sua iluminada presença.

Ressalto o papel dos meus orientadores Edmundo Albuquerque de Souza e Silva e Rosa Maria Meri Leão na minha formação acadêmica, profissional e pessoal; e ao empenho de ambos pelo ensino e pesquisa de qualidade no Brasil.

Por fim, registro o papel fundamental da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) no apoio deste e demais trabalhos voltados à pesquisa.

★

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## AVALIAÇÃO DE UM ALGORITMO ADAPTATIVO PARA PREVISÃO DE TAXAS DE PERDA DE PACOTES EM REDES DE COMPUTADORES

Hugo Hidequi Costa Sato

Dezembro/2007

Orientador: Edmundo Albuquerque de Souza e Silva

Programa: Engenharia de Sistemas e Computação

Em uma aplicação multimídia, ser capaz de prever a taxa de perda é um importante recurso para aumentar a qualidade de serviço eficientemente. Por exemplo, algoritmos de recuperação de perdas de pacotes (**FEC**) poderiam ser usados de forma adaptativa, enviando-se mais ou menos redundância, de acordo com a previsão. Viabilizando uma recuperação dos pacotes perdidos com menos uso de banda.

Neste trabalho, consideramos o problema de realizar a previsão da taxa de perda de pacotes em uma janela de tempo futura. Aqui, entende-se por previsão a estimativa de uma métrica da taxa de perda condicionada às observações passadas.

Para obter as previsões da taxa de perda, utilizaremos o algoritmo adaptativo de previsão proposto no trabalho de [Silveira e de Souza e Silva 2006]. Este algoritmo é responsável tanto por estimar quanto atualizar o modelo preditivo às variações do processo de perda ao longo do tempo e, em cada previsão, obter uma estimativa da taxa de perdas futura condicionada às observações das perdas passadas.

Desta forma, o objetivo desta dissertação é estender o trabalho iniciado em [Silveira e de Souza e Silva 2006] no sentido de: **1**– avaliar mais detalhadamente do que o estudo em [Silveira e de Souza e Silva 2006] a capacidade de previsão deste algoritmo adaptativo utilizando os modelos de Markov ocultos, modelos autorregressivos e alguns preditores simples; **2**– verificar a sensibilidade da eficácia do algoritmo de previsão em relação a seus parâmetros com os modelos e cenários aqui considerados; **3**– adotar experimentos mais gerais do que os previamente considerados.

Como resultado da avaliação, foi possível mostrar que modelo autorregressivo tem um desempenho razoável. Embora veremos que o modelo HMM é superior em vários cenários por representar melhor a variabilidade da taxa de perdas. Além disto, temos que algoritmo é robusto às variações de seus parâmetros.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ON AN EVALUATION OF AN ADAPTIVE ALGORITHM FOR  
PACKET LOSS FORECASTING OVER COMPUTER NETWORKS

Hugo Hidequi Costa Sato

December/2007

Advisor: Edmundo Albuquerque de Souza e Silva

Department: Systems Engineering and Computer Science

Loss rate prediction is an useful tool for improving the quality of service on multimedia application in computer networks. For instance, Forward Error Correction algorithms could be used in an adaptively fashion, sending more or less redundancy, in response to the prediction results. This would allow an efficient packet recovery in terms of throughput demand.

In this work, we consider the problem of packet loss rate prediction over a future time window. Here, prediction is an estimation of any loss rate related measure conditioned with recent observations of loss events.

The loss rate prediction is obtained with an adaptive prediction algorithm initially proposed in [Silveira e de Souza e Silva 2006]. This algorithm is responsible for estimating the predictive model and parameterize the model used dynamically according to the changes in loss process. At each prediction, the estimates the loss rate conditioned on the loss events in the past.

The main objective of this dissertation is to extend the work first initiated in [Silveira e de Souza e Silva 2006] as follows: **1**– to evaluate the prediction ability of this adaptive algorithm when the loss rate is modeled by hidden Markovian models, by autoregressive models, or by simple predictor models; **2**– to verify the sensitivity of the algorithm's efficacy in more details than the work aforementioned over its parameter values and the considered models and loss scenarios; **3**– to adopt distinct experimental configurations from the previously considered.

The results of this analysis show that the autoregressive model has reasonable performance. Nevertheless, the hidden Markovian model is more suited to represent the loss rate variability. Moreover, the adaptive algorithm is robust to variations of its parameter values.

# Sumário

<b>Resumo</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Conceitos teóricos</b>	<b>4</b>
2.1 Processos estocásticos . . . . .	4
2.1.1 Covariância e correlação cruzadas . . . . .	5
2.1.2 Autocovariância e autocorrelação . . . . .	7
2.1.3 Entropia . . . . .	8
2.2 Modelos de Markov ocultos . . . . .	10
2.2.1 Estimação de parâmetros . . . . .	11
2.2.2 Estimando o número de estados do modelo . . . . .	16
2.3 Modelos autorregressivos . . . . .	17
2.3.1 Estimação de parâmetros . . . . .	18
2.4 Problema de previsão . . . . .	19

2.4.1	Preditores lineares . . . . .	19
2.4.2	Análise de erro de dois preditores simples . . . . .	21
<b>3</b>	<b>Trabalhos relacionados e o algoritmo adaptativo de previsão</b>	<b>23</b>
3.1	Trabalhos relacionados . . . . .	23
3.2	Medida da previsão . . . . .	25
3.3	Algoritmo adaptativo para previsão da taxa de perda . . . . .	28
3.3.1	Algoritmo adaptativo com modelos de Markov ocultos . . . . .	29
3.3.2	Algoritmo adaptativo com modelos autorregressivos . . . . .	30
<b>4</b>	<b>Resultados experimentais</b>	<b>32</b>
4.1	Capacidade de previsão do algoritmo adaptativo . . . . .	33
4.2	Erro de estimação do modelo HMM . . . . .	35
4.3	Resultados usando modelos analíticos . . . . .	37
4.4	Cenários com <i>traces</i> . . . . .	43
4.5	Cenários simulados . . . . .	51
4.6	Resumo dos experimentos . . . . .	60
<b>5</b>	<b>Considerações finais e trabalhos futuros</b>	<b>62</b>
<b>A</b>	<b>Conceitos teóricos adicionais</b>	<b>64</b>
A.1	Erro médio quadrático e o preditor ótimo . . . . .	64
A.2	Modelos de Gilbert-Elliot . . . . .	65
A.3	Distribuição da média amostral . . . . .	67



A.4	Distribuição da função de autocorrelação amostral - SACF . . . . .	68
A.5	Distribuição multinormal . . . . .	69
<b>Referências Bibliográficas</b>		<b>70</b>

# Lista de Figuras

1.1	Análise preditiva . . . . .	2
2.1	Procedimento <i>forward</i> . . . . .	13
2.2	Procedimento <i>backward</i> . . . . .	14
2.3	Relação entre os erros de 3 preditores simples . . . . .	21
3.1	Modelo HMM hierárquico considerado nesta dissertação . . . . .	26
3.2	Taxa de perda $R_t^F$ . . . . .	26
3.3	Janela de previsão $F$ hipotética . . . . .	27
3.4	Coeficiente de variação da taxa de perda para diferentes valores de $F$ . . . . .	27
3.5	Algoritmo adaptativo em função dos parâmetros de estimação $(T, \tau)$ e de previsão $(H, F)$ . . . . .	28
3.6	Relação entre um evento de previsão hipotético com modelo HMM e modelo AR(3) . . . . .	31
4.1	Preditor constantemente nulo em um processo com picos de perdas . . . . .	34
4.2	Exemplo de uma rajada média amostral . . . . .	35

4.3	Autocorrelação da taxa de perda ( $F = 20$ ) para um modelo de perdas de Gilbert-Elliot simplificado com a fração de perda de 10% e rajadas de perdas média de 20 pacotes . . . . .	37
4.4	Comparação entre o erro médio quadrático da média amostral e o erro médio quadrático da média do modelo HMM estimado para $p = 0.005$ e $q = 0.05$ . . . . .	39
4.5	MSE dos modelos HMM, AR, média e replicador . . . . .	40
4.6	Relação entre a previsão e o histórico de previsão $H$ . . . . .	41
4.7	Previsão de um modelo de perdas de Gilbert-Elliot simplificado . . .	42
4.8	Previsão de um modelo de perdas de Gilbert-Elliot simplificado . . .	42
4.9	Taxa de perda de três <i>traces</i> selecionados para uma janela de previsão $F = 400$ milissegundos . . . . .	44
4.10	Características da taxa de perda <i>traces</i> selecionados . . . . .	45
4.11	Atraso das observações em um dado evento de previsão . . . . .	46
4.12	MSE do mecanismo de previsão adaptativa variando o histórico da previsão $H$ para o <i>trace</i> da Figura 4.9(a) . . . . .	47
4.13	Previsão em uma medição entre UMD e UFRJ . . . . .	47
4.14	Previsão em uma medição entre UMass e UFRJ . . . . .	48
4.15	Previsão em uma medição entre UFMG e UFRJ . . . . .	49
4.16	MSE do mecanismo de previsão adaptativa variando o intervalo de estimação da previsão $\tau$ . . . . .	50
4.17	Topologia geral da simulação . . . . .	51
4.18	Taxa de perda e autocorrelação em uma ambiente simulado com tráfego TCP . . . . .	52

4.19	Cadeia com estrutura particular <i>coxian</i> . . . . .	53
4.20	MSE do mecanismo de previsão adaptativa variando o histórico da previsão $H$ com uma janela de previsão de $400ms$ em um ambiente simulado com tráfego TCP . . . . .	54
4.21	Previsão em um ambiente simulado com tráfego TCP . . . . .	54
4.22	Autocorrelação da taxa de perdas em um ambiente simulado com tráfego MMPP . . . . .	55
4.23	Intensidade de tráfego em quatro diferentes escalas de tempo para um processo Poisson e um processo MMPP ambos com mesma média . . . . .	56
4.24	Convergência da HMM de 5 estados com tráfego MMPP . . . . .	57
4.25	MSE do mecanismo de previsão adaptativa variando o histórico da previsão $H$ com uma janela de previsão de $400ms$ em um ambiente simulado com tráfego MMPP . . . . .	57
4.26	Taxa de perda e previsão em um cenário simulado com tráfego de fundo MMPP . . . . .	59
A.1	Modelo de Gilbert-Elliot de tempo discreto . . . . .	66
A.2	Modelo de <b>Gilbert-Elliot simplificado</b> de tempo discreto . . . . .	67

# Lista de Tabelas

4.1	Comparação de medidas de desempenho para uma janela de previsão $F = 20$ , $H = 20$ pacotes ( $p = 1$ ) e $T = 1000$ pacotes . . . . .	41
4.2	Comparação de medidas de desempenho para uma janela de previsão $F = 400$ milissegundos no <i>trace</i> da Figura 4.9(a), $H = 2$ segundos e $T = 2$ minutos . . . . .	47
4.3	Comparação de medidas de desempenho para uma janela de previsão $F = 400$ milissegundos no <i>trace</i> da Figura 4.9(b), $H = 0.8$ segundos e $T = 10$ segundos . . . . .	48
4.4	Comparação de medidas de desempenho para uma janela de previsão $F = 400$ milissegundos no <i>trace</i> da Figura 4.9(c), $H = 800$ milissegun- dos e $T = 1$ minuto . . . . .	49
4.5	Comparação de medidas de desempenho para uma janela de previsão $F = 400$ milissegundos em um ambiente simulado com tráfego TCP .	53
4.6	Comparação de medidas de desempenho para uma janela de previsão $F = 400$ milissegundos, $H = 400$ milissegundos ( $p=1$ ) e $T = 1$ minuto em um ambiente simulado com tráfego MMPP . . . . .	58
4.7	Relação de parâmetros e medidas de desempenho para uma janela de previsão $F = 400$ milissegundos em todos os cenários . . . . .	61

# Capítulo 1

## Introdução

Nas aplicações multimídias em redes de computadores **IP**, a perda de pacotes de dados é uma das causas da degradação da qualidade de serviço prestado. Assim, uma das imposições à melhoria desta é a maneira como o problema da perda de pacotes transmitidos é tratado. Uma estratégia possível é prever a taxa de perda em um intervalo de tempo futuro, permitindo uma adaptação pro-ativa por parte da aplicação.

Por exemplo, existem aplicações de voz sobre **IP** que utilizam diferentes esquemas de redundância (**FEC**) para mascarar a perda de pacotes, de acordo com a previsão da taxa de perda futura, e.g. [Silveira et al. 2006, Duarte et al. 2003, Bolot et al. 1999].

Outro exemplo da aplicação da previsão da taxa de perda é na escolha por qual das rotas enviar o fluxo de dados em redes com roteamento por múltiplos caminhos, e.g. redes *overlay*. Em [Tao e Guerin 2004], os autores apresentam um mecanismo para prever a taxa de perda de um dado caminho. A partir desta estimativa, o desempenho que aplicação teria sobre aquela rota é avaliado.

Também poderíamos utilizar a previsão da taxa perdas para ajustar a taxa de transmissão em um mecanismo de controle de fluxo. Se a aplicação tem uma esti-

mativa da taxa de perda futura, podemos controlar seu fluxo de dados a partir de modelos analíticos do *throughput* do TCP. O objetivo deste “controle preditivo” é ter um fluxo de dados TCP-*friendly* sem a variabilidade observada naquele tipo de fluxo.

Nesta dissertação, consideramos o problema de previsão da taxa de perda. Uma motivação inicial era o desenvolvimento de um algoritmo de controle de fluxo preditivo conforme descrito anteriormente. No entanto, este objetivo específico será estudado em um trabalho futuro, uma vez que o estudo detalhado realizado é por si só uma tarefa relevante.

Abordaremos o problema de previsão através da metodologia conhecida por análise preditiva. A Figura 1.1 ilustra os princípios gerais desta análise. Em resumo, temos um modelo preditivo para representar o processo de perdas e, a partir deste, obteremos as previsões da taxa de perda. Para avaliar a acurácia do resultado da previsão, utiliza-se um caminho amostral do processo de perdas.

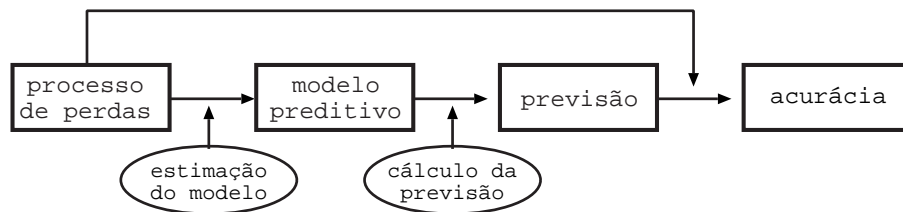


Figura 1.1: Análise preditiva

Para obter as previsões da taxa de perda, utilizaremos o algoritmo adaptativo de previsão proposto no trabalho de [Silveira e de Souza e Silva 2006]. Este algoritmo é responsável tanto por estimar quanto parametrizar dinamicamente o modelo preditivo de acordo com as variações do processo de perda ao longo do tempo e, em cada previsão, obter uma estimativa da taxa de perdas futura condicionada às observações das perdas passadas.

Ainda naquele trabalho, os autores utilizam um modelo de Markov oculto hierárquico para descrever as características do processo de perdas. A acurácia deste modelo foi comparada com a de outros modelos de Markov ocultos. No entanto,

uma avaliação comparativa deste modelo de Markov oculto com os modelos autor-regressivos, freqüentemente utilizados em análise preditiva, não foi extensivamente realizada no passado. Além disto, havia a necessidade de uma análise sistemática a respeito da sensibilidade dos parâmetros do algoritmo adaptativo na acurácia da previsão.

Desta forma, o objetivo desta dissertação é estender o trabalho iniciado em [Silveira e de Souza e Silva 2006] no sentido de: **1**– avaliar com maiores detalhes a capacidade de previsão deste algoritmo adaptativo utilizando os modelos de Markov ocultos, modelos autorregressivos, e alguns preditores simples; **2**– verificar a eficácia do algoritmo de previsão em relação a seus parâmetros com os modelos e cenários aqui considerados; **3**– utilizar cenários de experimentação mais diversificados dos aqueles previamente considerados.

Uma contribuição fundamental deste trabalho é o desenvolvimento de uma metodologia para avaliação da acurácia das previsões baseada em medidas quantitativas. Como resultado da avaliação, foi possível mostrar que o algoritmo é robusto às variações de seus parâmetros.

Além disto, propomos a utilização de um modelo preditivo autorregressivo comparativamente ao modelo HMM. Embora o modelo AR tenha um desempenho razoável, veremos que o modelo HMM é superior em vários cenários e representa melhor a variabilidade da taxa de perdas.

Este trabalho também contribui com uma revisão de diversas técnicas matemáticas aplicadas nos problemas relacionados direta ou indiretamente com análise preditiva.

O texto restante está organizado da seguinte forma: no próximo capítulo, iremos descrever os principais conceitos teóricos utilizados ao longo da dissertação. No terceiro capítulo, revisaremos os trabalhos relacionados e detalharemos o algoritmo adaptativo mencionado anteriormente. Então, no quarto capítulo, descreveremos os principais resultados experimentais; finalizando no capítulo seguinte.



# Capítulo 2

## Conceitos teóricos

Neste capítulo, apresentamos os conceitos teóricos relevantes para este trabalho. Começaremos definindo um processo estocástico e suas estatísticas relevantes para os próximos capítulos. Em seguida, abordaremos os dois modelos preditivos utilizados no algoritmo adaptativo: os modelos de Markov ocultos (**HMM**) e modelos autorregressivos (**AR**), conforme mencionado no Capítulo 1. Para cada modelo respectivamente, apresentaremos os métodos utilizados na estimação de seus parâmetros. Ao final do capítulo, formalizamos o problema de previsão.

Uma abordagem completa sobre modelos de Markov ocultos e modelos autorregressivos pode ser encontrada respectivamente em [Rabiner 1989, Elliot et al. 1995] e [Box et al. 1994, Brockwell e Davis 2002].

### 2.1 Processos estocásticos

Para representar analiticamente o processo de perdas, utilizamos a noção de processo estocástico. Formalmente, um processo estocástico  $X = \{X_t : t \in \mathbf{T}\}$ , denotado por  $\{X_t\}$ , é um conjunto de variáveis aleatórias [Ross 1982]. Caracterizamos  $\{X_t\}$  por três elementos, a seguir

1. Espaço amostral  $S$ . Seja  $x_t$  uma realização qualquer de um elemento de  $X_t$ ,  $x_t \in S, \quad \forall t \in \mathbf{T}$ .
2. Conjunto de índices  $\mathbf{T}$ . Se  $\mathbf{T}$  é um conjunto contável,  $\{X_t\}$  é dito um processo estocástico discreto e se  $\mathbf{T}$  é contínuo, chamamos de processo estocástico contínuo.
3. Função de distribuição  $f_t(x) = P[X_t = x], \quad \forall t \in \mathbf{T}$ , ou função de distribuição cumulativa  $F_t(x) = P[X_t \leq x], \quad \forall t \in \mathbf{T}$ .

Não raro, a função de distribuição  $f_t(x)$  não é propriamente descrita. Em contrapartida, caracterizam-a por uma propriedade mais geral. Por exemplo, um processo estocástico Markoviano é caracterizado pela sua ausência de memória, isto é

$$P[X_t = x_t | X_{t-1} = x_{t-1} \cdots X_1 = x_1] = P[X_t = x_t | X_{t-1} = x_{t-1}] = a_{x_t, x_{t-1}}. \quad (2.1)$$

Quando o conjunto  $\mathbf{T}$  é interpretado como tempo, um processo estocástico é comumente denominado série temporal, enquanto interpretado como espaço é denominado campo randômico. Como nosso interesse é analisar e prever um processo ao longo do tempo, utilizaremos as expressões série temporal, ou **série**, e processo estocástico, ou **processo**, sem distinções\*.

A seguir descreveremos algumas estatísticas importantes tanto para descrever características do processo de perdas quanto para verificar a acurácia da previsão.

### 2.1.1 Covariância e correlação cruzadas

Dado dois processos estocásticos  $\{X_t\}$  e  $\{Y_t\}$ , um método para estimarmos o grau como estas duas séries estão correlacionadas é através da covariância cruzada<sup>†</sup> [Bendat e Piersol 1971].

---

\*Neste sentido, também utilizaremos o termo **modelo** para expressar o conceito de processo estocástico.

<sup>†</sup>O adjetivo cruzado é usado para enfatizar a que temos a covariância entre dois processos estocásticos distintos e a noção de matriz de covariância de um vetor de variáveis aleatórias.

A função de covariância cruzada é escrita como

$$\gamma_{XY}(t) = Cov(X_t, Y_t) = E[(X_t - \mu_X(t))(Y_t - \mu_Y(t))], \quad (2.2)$$

para  $t \in \mathbf{T}$ ,  $\mu_X(t) = E[X_t]$  e  $\mu_Y(t) = E[Y_t]$ .

Através da definição, vemos que a medida que  $\{X_t\}$  e  $\{Y_t\}$  variam conjuntamente em torno das respectivas médias o valor absoluto de  $\gamma_{XY}(t)$  aumenta.

A medida normalizada em  $[-1, 1]$  é a correlação cruzada, dada por

$$\rho_{XY}(t) = \frac{Cov(X_t, Y_t)}{\sqrt{Var[X_t]} \cdot \sqrt{Var[Y_t]}}. \quad (2.3)$$

Quando o valor de  $|\rho_{XY}(t)|$  tende a 1 (0) mais (menos) correlacionadas são ambas as séries. Se  $\rho_{XY}(t)$  tende a  $-1$ , temos uma alta correlação mas em sentidos opostos em relação à média.

A correlação cruzada é uma das medida da acurácia da previsão. Com ela, temos a noção se a “variabilidade instantânea” da série formada pelas previsões estará associada com uma variabilidade do processo de perdas.

## Estimadores

Seja  $\mathbf{x} = x_1 \cdots x_n$  e  $\mathbf{y} = y_1, \cdots, y_n$  uma realização do processo estocástico  $\{X_t\}$  e  $\{Y_t\}$ , também chamado de caminho amostral. Assumindo que  $\gamma_{XY}$  seja independente de  $t$ , o estimador para a covariância cruzada é dada por

$$\hat{\gamma}_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)}{n - 1}, \quad (2.4)$$

onde  $\bar{\mu}_x$  é média amostral definida pela Equação A.6. O estimador da correlação cruzada é dado por

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)}{\sqrt{\sum_{i=1}^n (x_i - \bar{\mu}_x)^2 \sum_{i=1}^n (y_i - \bar{\mu}_y)^2}}. \quad (2.5)$$

### 2.1.2 Autocovariância e autocorrelação

Se desejamos medir a correlação temporal de um processo  $\{X_t\}$ , utilizamos a função de covariância definida na Equação 2.2 como

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]. \quad (2.6)$$

É imediato observar que  $\gamma_X(r, r) = \text{Var}[X_r] = \sigma_X^2(r)$ .

Dizemos que  $\{X_t\}$  é estacionário de segunda ordem ou simplesmente **estacionário** se:

1.  $E[X_t] = \mu_X$  é independente de  $t$ .
2.  $\gamma_X(t + h, t)$  é independente de  $t$  para qualquer valor de  $h$ .

Sendo  $\{X_t\}$  um processo estocástico **estacionário**, reescrevemos a Expressão 2.6 e obtemos a função de autocovariância (ACVF) de  $\{X_t\}$  para um lag  $h$ :

$$\gamma_X(h) = \gamma_X(t + h, t) = \gamma_X(h, 0) = \text{Cov}(X_h, X_0). \quad (2.7)$$

Como medida normalizada entre  $[-1, 1]$ , temos a função de autocorrelação (ACF) de  $\{X_t\}$ :

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}. \quad (2.8)$$

A seguir, definiremos uma classe importante de processos estocásticos em termos da sua função de autocorrelação:

**Definição 2.1.** *Seja  $\{X_t\}$  um processo estocástico **descorrelacionado**, i.e.  $\rho(h) = 0$ ,  $|h| > 0$ , com média 0 e variância  $\sigma^2$ . Tal processo é denominado ruído branco ou simplesmente WN (white noise) e denotamos por  $X_t \sim \text{WN}(0, \sigma^2)$ .*

Este processo estocástico será útil ao caracterizarmos processos lineares, como veremos na Seção 2.3.

## Estimadores

Um processo estacionário  $\{X_t\}$  pode ser caracterizado pelos descritores de segunda ordem através de  $\mu_X = E[X_t]$  e sua função de autocovariância  $\gamma_X(\cdot)$ . Portanto, estimar essas quantidades é de fundamental importância.

Seja  $\mathbf{x} = x_1 \cdots x_n$  uma realização do processo estocástico  $\{X_t\}$ . A função de autocovariância amostral (SACVF) é

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (x_{i+|h|} - \bar{\mu}_x)(x_i - \bar{\mu}_x), \quad -n < h < n. \quad (2.9)$$

Analogamente, a função de autocorrelação (SACF) amostral é

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n. \quad (2.10)$$

### 2.1.3 Entropia

Através da noção de entropia podemos quantificar a informação contida em um processo estocástico. Esta medida nos será útil para comparar a informação em um caminho amostral, de um processo estocástico desconhecido, e em um processo estocástico arbitrado [Ziv e Merhav 1992].

Seja  $X_1, \dots, X_n$  uma sequência de variáveis aleatórias, a entropia conjunta desta sequência é definida como

$$H(X_1, \dots, X_n) = - \sum_{(x_1, \dots, x_n) \in S^n} p(x_1, \dots, x_n) \log_2 p(x_1, \dots, x_n) \quad (2.11)$$

Então, definimos a taxa de entropia de um processo estocástico por

$$H(\mathcal{X}) = - \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n), \quad (2.12)$$

a taxa de entropia nos diz como a informação do processo estocástico  $\{X_t\}$  cresce ao aumentarmos o número de elementos considerados.

## Estimadores

Para estimar a taxa de entropia de um processo estocástico, nos apoiaremos no Teorema 2.1, a ser apresentado, do limite de compressão de um conjunto de amostras do processo estocástico  $\{X_t\}$ .

Antes de enunciar o teorema, precisamos definir a noção de codificação:

**Definição 2.2.** Uma codificação  $C_n$  é um mapeamento de  $S^n$ , o espaço amostral de um conjunto de  $n$  símbolos representados pelas variáveis aleatórias  $X_1, \dots, X_n$ , para  $D^*$ , o conjunto de códigos finitos criados a partir de um alfabeto  $D$ -ário.  $C(\mathbf{x})$  denota o código correspondente a realização  $\mathbf{x} = x_1 \cdots x_n$  enquanto  $l(\mathbf{x})$  denota o seu respectivo tamanho em bits.

Com isto, temos a seguinte definição de código médio,

**Definição 2.3.** O código médio por símbolo,  $L_n$ , de uma codificação  $C_n$  é dado por

$$L_n = \frac{1}{n} \sum_{\mathbf{x} \in S^n} p(\mathbf{x}) l(\mathbf{x}) = \frac{E[l(\mathbf{x})]}{n} \quad (2.13)$$

Assim, o teorema do limite de compressão é

**Teorema 2.1.** O código médio mínimo,  $L_n^*$ , satisfaz as seguintes inequações

$$\frac{H(X_1, X_2, \dots, X_n)}{n} \leq L_n^* < \frac{H(X_1, X_2, \dots, X_n)}{n} + \frac{1}{n} \quad (2.14)$$

Se  $X_t$  é um processo estocástico estacionário, então

$$L_n^* \rightarrow H(\mathcal{X}) \quad (2.15)$$

Desta forma, a taxa de entropia é um limite inferior para a compressão de uma sequência de amostras de um dado processo estocástico. Se temos um conjunto suficientemente grande amostras de  $X_t$ , ao utilizarmos um compressor ideal seremos capazes de estimar  $H(\mathcal{X})$  apoiado no teorema acima. A prova deste teorema pode ser encontrada [Cover e Thomas 2006].

## 2.2 Modelos de Markov ocultos

Modelos de Markov ocultos, ou simplesmente **HMM** (*Hidden Markovian Models*), têm sido utilizados em inúmeras áreas, tais como reconhecimento de padrões, bioinformática e processamento de sinais, para citar algumas.

Neste trabalho, utilizamos um modelo de Markov oculto como modelo preditivo de referência no algoritmo de previsão adaptativo. Adotamos o modelo proposto em [Silveira e de Souza e Silva 2006], conforme mencionado no Capítulo 1. Suas características específicas serão descritas no capítulo seguinte, ao detalharmos os trabalhos relacionados. Nesta seção, descrevemos os aspectos teóricos de um HMM geral, baseado no trabalho de [Rabiner 1989].

Assim, considere um processo estocástico Markoviano  $\{Y_t\}$  (ver Equação 2.1) de tempo discreto com  $N$  estados e um processo estocástico de observações  $\{X_t\}$  com  $M$  estados. Um modelo de Markov oculto é definido como uma extensão do processo  $Y_t$  onde o estado observado  $x_t$  é uma **função probabilística** do estado  $y_t$ . Ou seja, os processos estocásticos  $\{X_t\}$  e  $\{Y_t\}$  são dependentes entre si. Neste modelo resultante,  $\{Y_t\}$  deixa de ser observado diretamente, daí o termo oculto.

Todo HMM é definido pelos parâmetros:

1. Número de estados  $N$  do processo Markoviano  $\{Y_t\}$ .
2. Número de observações  $M$  do processo  $\{X_t\}$ .
3. Vetor  $\pi = [\pi_i]$  de distribuição do estado inicial de  $\{Y_t\}$ , onde  $\pi_i = P[Y_0 = i]$ ,  $1 \leq i \leq N$ .
4. Matriz  $\mathbf{A} = [a_{ij}]$  de dimensão  $N \times N$  das probabilidades de transição entre estados de  $\{Y_t\}$ , onde

$$a_{ij} = P[Y_{t+1} = i | Y_t = j] \quad 1 \leq i, j \leq N$$

5. Matriz  $\mathbf{B} = [b_{ij}]$  de dimensão  $N \times M$  de distribuição probabilística da observação  $j$  dado o oculto estado  $i$ , onde

$$b_{ij} = P[X_t = j | Y_t = i], \quad 1 \leq i \leq N, \quad 1 \leq j \leq M$$

sucintamente, denotamos o HMM pela tripla  $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ .

### 2.2.1 Estimação de parâmetros

Dada a definição do modelo preditivo HMM, na estimação de seus parâmetros  $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$  teremos o modelo para representar uma instância particular do processo considerado, no nosso caso, o processo de perdas.

A estimação é dada pela maximização da função de verossimilhança dos dados, i.e., a maximização da probabilidade das observações  $\mathbf{x} = x_1, \dots, x_n$  terem sido emitidas pelo HMM. Denotamos a função de verossimilhança por  $P[\mathbf{x}|\lambda]$ . Este problema é resolvido pelo algoritmo iterativo de **Baum-Welch**. Antes de descrever tal algoritmo, consideremos o cômputo de  $P[\mathbf{x}|\lambda]$ .

A forma mais imediata de calcularmos  $P[\mathbf{x}|\lambda]$  é enumerar todas as seqüências de estados distintas e viáveis, no sentido que haja um caminho de transições entre os estados dada a matriz  $\mathbf{A} = [a_{ij}]$ , de tamanho  $n$  e somar suas probabilidades, pelo teorema da probabilidade de total:

$$\begin{aligned} P(\mathbf{x}|\lambda) &= \sum_{(y_1, \dots, y_n)} P(\mathbf{x}, (y_1, \dots, y_n) | \lambda) \\ &= \sum_{(y_1, \dots, y_n)} P(\mathbf{x} | (y_1, \dots, y_n), \lambda) \cdot P((y_1, \dots, y_n) | \lambda). \end{aligned} \quad (2.16)$$

Utilizando a definição de HMM acima e independência causal entre as observações dado a seqüência de estados ([Rabiner 1989]),

$$P(\mathbf{x} | (y_1, \dots, y_n), \lambda) = b_{y_1, x_1} \cdot b_{y_2, x_2} \cdots b_{y_n, x_n}, \quad (2.17)$$

organizando os fatores de maneira conveniente, temos

$$P(\mathbf{x}|\lambda) = \sum_{(y_1, \dots, y_n)} \pi_{y_1} b_{y_1, x_1} \cdot a_{y_1, y_2} b_{y_2, x_2} \cdots a_{y_{n-1}, y_n} b_{y_n, x_n}. \quad (2.18)$$



Apesar da simplicidade da derivação acima, este método possui uma complexidade não polinomial em relação a  $n$ , tornando-o inviável. Felizmente, existe um método baseado em indução matemática chamado *forward-backward*, que é capaz de efetuar este cômputo de forma eficiente.

Seja a variável *forward*  $\alpha_t(i)$  a variável de indução:

$$\alpha_t(i) = P(x_1, x_2 \cdots x_t, y_t = i | \lambda), \quad (2.19)$$

que representa um sub-problema mais simples: obter a probabilidade da subsequência de observações,  $x_1, x_2 \cdots x_t$ , tendo o estado  $i$  na  $t$ -ésima transição, dado o modelo  $\lambda$ . O cálculo desta variável é efetuado da seguinte forma:

1. Inicialização:

$$\alpha_1(i) = \pi_i \cdot b_{i,x_1}, \quad 1 \leq i \leq N, \quad (2.20a)$$

pela própria definição de  $\alpha_t(i)$

2. Indução:

$$\begin{aligned} \alpha_{t+1}(i) &= \left[ \sum_{j=1}^N P(x_1, \cdots x_t, y_t = j | \lambda) P[y_{t+1} = i | y_t = j] \right] P[x_{t+1} | y_{t+1} = i] \\ &= \left[ \sum_{j=1}^N \alpha_t(j) a_{ji} \right] b_{i,x_{t+1}} \end{aligned} \quad (2.20b)$$

Onde  $1 \leq t \leq n - 1$  e  $1 \leq i \leq N$ .

3. Término:

$$P(\mathbf{x} | \lambda) = \sum_{i=1}^N \alpha_n(i) \quad (2.20c)$$

A Figura 2.1 ilustra a árvore de construção do algoritmo, da esquerda para a direita. Analisando a indução, temos que a complexidade do procedimento *forward* é de  $O(N^2n)$  operações, sendo  $N(N+1)(n-1) + N$  multiplicações e  $N(N-1)(n-1)$  adições precisamente. A Equação 2.20c nos fornece uma maneira eficiente de obter  $P(\mathbf{x} | \lambda)$ . Entretanto, uma variável complementar denominada *backward* será

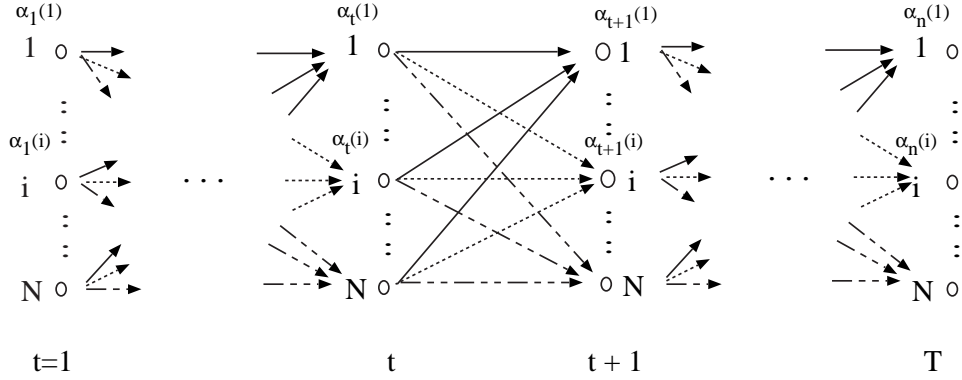


Figura 2.1: Procedimento *forward*

utilizada posteriormente. Considere, então, a variável *backward*  $\beta_t(i)$  definida como:

$$\beta_t(i) = P(x_{t+1}, x_{t+2} \cdots x_n | y_t = i, \lambda), \quad (2.21)$$

representando a probabilidade da sequência parcial de observações de  $t + 1$  até  $n$  dado o estado  $i$  no tempo  $t$  e o modelo  $\lambda$ . A variável  $\beta_t(i)$  pode ser calculada com a indução:

1. Inicialização:

$$\beta_n(i) = 1, \quad 1 \leq i \leq N \quad (2.22a)$$

Resulta da definição de  $\beta_t(i)$ .

2. Indução:

$$\begin{aligned} \beta_t(i) &= \sum_{j=1}^N P(y_{t+1} = j | y_t = i) P(x_{t+1} | y_{t+1} = j) P(x_{t+2}, \cdots, x_n | y_{t+1} = j) \\ &= \sum_{j=1}^N a_{ij} b_{j, x_{t+1}} \beta_{t+1}(j) \end{aligned} \quad (2.22b)$$

Onde  $t = n - 1, n - 2, \cdots, 1$  e  $1 \leq i \leq N$ .

A Figura 2.2 ilustra a árvore de construção do algoritmo da direita para a esquerda. Pelo procedimento acima, podemos observar que o método *backward* possui mesma complexidade do *forward*.

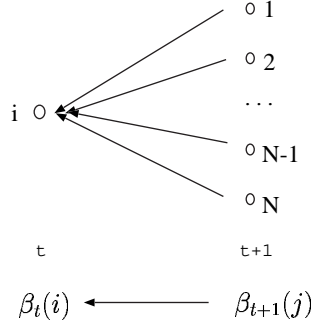


Figura 2.2: Procedimento *backward*

As seguintes identidades podem ser estabelecidas:

$$\gamma_t(i) = P(Y_t = i | \mathbf{X}, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}, \quad (2.23a)$$

$$\xi_t(i, j) = P(Y_t = i, Y_{t+1} = j | \mathbf{X}, \lambda) = \frac{\alpha_t(i)a_{ij}b_{jx_{t+1}}\beta_{t+1}(j)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}. \quad (2.23b)$$

Agora podemos voltar ao problema inicial que consistia, em dada uma seqüência  $\mathbf{x}$  de observações de tamanho  $n$ , obter  $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$  de forma a maximizar a função objetivo  $P(\mathbf{x}|\lambda)$ . Como não existe uma expressão analítica que maximize a verossimilhança, uma solução possível é obtida pelo método iterativo de Baum-Welch – que não garante a convergência à solução ótima global. Entretanto, a cada nova iteração, a função objetivo é incrementada, garantindo a convergência do modelo para um máximo local qualquer. O algoritmo consiste dos seguintes passos:

1. O primeiro passo é escolher os valores iniciais para os parâmetros  $\pi$ ,  $\mathbf{A}$  e  $\mathbf{B}$  do modelo.
2. O segundo passo é o cálculo das variáveis *forward*  $\alpha_t(i)$  e *backward*  $\beta_t(i)$ . Com esses valores pode-se avançar para o próximo passo, que é o cálculo do número esperado de visitas ao estado  $i$  no tempo  $t$ ,  $\gamma_t(i)$ , e o número esperado de transições de  $i$  para  $j$ ,  $\xi_t(i, j)$ .
3. Então reestimamos  $\lambda$ . Os parâmetros do novo modelo ( $\bar{\lambda}$ ) são calculados pelas expressões abaixo:

$$\begin{aligned}
\bar{\pi}_i &= \text{número esperado de visitas ao estado } i \text{ no instante de tempo } (t = 1) \\
&= \gamma_1(t) \\
\bar{a}_{ij} &= \frac{\text{número esperado de transições do estado } i \text{ para o estado } j}{\text{número esperado de transições a partir de } i} \\
&= \frac{\sum_{t=1}^{n-1} \xi_t(i, j)}{\sum_{t=1}^{n-1} \gamma_t(i)} \\
\bar{b}_{jk} &= \frac{\text{número esperado de vezes no estado } j \text{ emitindo o símbolo } k}{\text{número esperado de vezes no estado } j} \\
&= \frac{\sum_{t=1}^n \gamma_t(j)}{\sum_{t=1}^n \gamma_t(j)} \quad (2.24)
\end{aligned}$$

4. De posse dos novos parâmetros, a verossimilhança do modelo reestimado é calculada. Esta probabilidade é comparada a do modelo antigo e se o ganho relativo  $(P(\mathbf{x}|\bar{\lambda}) - P(\mathbf{x}|\lambda))$  for pequeno pode-se interromper o algoritmo. Também é possível encerrar pelo número de iterações, uma vez que o tempo para atingir um ganho relativamente pequeno pode ser relativamente grande, principalmente se este algoritmo estiver sendo usado em uma aplicação de tempo real.

As fórmulas de reestimação no passo 3 são, na verdade, um caso particular do algoritmo EM no qual o passo E (*Expectation*) é o cálculo da função auxiliar  $Q(\lambda, \bar{\lambda})$  e o passo M (*Maximization*) é a maximização de  $Q(\lambda, \bar{\lambda})$  em relação aos novos parâmetros  $(\bar{\lambda})$  dado os atuais  $(\lambda)$ .

Outra abordagem possível, é resolver este problema através dos métodos de gradiente para otimização de uma função objetivo. De fato, estes métodos foram comparados com esta instância do algoritmo EM e mostraram resultados semelhantes [Rabiner 1989].

Por ser um método iterativo com condição de parada sobre um máximo local, a escolha dos valores iniciais é fundamental. Entretanto, não existe um método para tal escolha, sendo a escolha aleatória a estratégia mais adotada em problemas práticos.

Em termos de complexidade, a cada iteração temos que tabular os valores de *forward*  $\alpha_t(i)$  e da variável *backward*  $\beta_t(i)$ , para  $1 \leq t \leq n$  e  $1 \leq i \leq N$ , com  $O(N^2n)$  operações no total. Então, para o cálculo de  $\pi = [\pi_i]$ ,  $\mathbf{A} = [a_{ij}]$  e  $\mathbf{B} = [b_{ij}]$ , haverão  $O(nN)$ ,  $O(nN^2)$  e  $O(nNM)$  operações respectivamente. Com uma complexidade total por iteração de  $O(nN(1 + N + M))$ . Em nossos experimentos, notamos que o algoritmo acima raramente ultrapassa 100 iterações, levando em média 50 iterações, para uma condição de parada  $P(\mathbf{x}|\bar{\lambda}) - P(\mathbf{x}|\lambda) < 10^{-6}$ . Nas atuais máquinas, o tempo de computação é ínfimo.

## 2.2.2 Estimando o número de estados do modelo

Existem situações, ao se modelar uma série  $\{X_t\}$  através de modelos de Markov ocultos, em que o número de estados ocultos não possui uma relação imediata com o processo real, o que nos leva a algum método de estimação. Adotaremos a mesma metodologia utilizada por [Salamatian e Vaton 2001] que consiste em aproximar a taxa de entropia do HMM à do processo  $\{X_t\}$ .

Assim, temos que calcular a taxa de entropia  $H(\mathcal{X}|\lambda^*)$  do HMM, onde  $\lambda^*$  é o HMM de maior verossimilhança para uma realização  $\mathbf{x} = x_1, \dots, x_n$  de  $\{X_t\}$ , e a taxa entropia  $H(\mathcal{X})$  do processo estocástico  $\{X_t\}$ . Então, estimaremos o número de estados  $N$  do HMM através de

$$N^* = \arg \min_j \{j \text{ estados ocultos} : |H(\mathcal{X}|\lambda^*) - H(\mathcal{X})| < \epsilon\}$$

O valor de  $H(\mathcal{X}|\lambda^*)$  é dado pela **propriedade da equipartição assintótica** [Cover e Thomas 2006], i.e.  $P(\mathbf{x}|\lambda^*) = 2^{-nH(\mathcal{X}|\lambda^*)}$ . Enquanto a estimação de  $H(\mathcal{X})$  é feita pelo código de compressão ótimo, conforme o Teorema 2.1. Assim, o estimador

é reescrito como

$$N^* = \arg \min_j \left\{ j \text{ estados ocultos} : \left| -\frac{1}{n} \log_2 P(\mathbf{x}|\lambda^*) - L^*(\mathbf{x}) \right| < \epsilon \right\}$$

Em [Merhav et al. 1989], os autores mostram que este estimador é consistente e assintoticamente ótimo com um número suficientemente grande de amostras do processo  $\{X_t\}$ . No Capítulo 4, veremos que para processos especificamente escolhidos o número de estados é estimado corretamente.

## 2.3 Modelos autorregressivos

Consideraremos agora a família de processos lineares. Tais processos constituem uma importante classe de processos estocásticos. De fato, todo processo estacionário pode ser descrito, em termos de estatísticas de segunda ordem, por um processo linear [Brockwell e Davis 2002]. Formalmente, os processos lineares são definidos por

**Definição 2.4.** *Uma série temporal  $\{X_t\}$  é um processo linear se*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (2.25)$$

para todo  $t$ ,  $Z_t \sim WN(0, \sigma^2)$ . Como condição de convergência do somatório,  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ .

Uma sub-classe de processos lineares muito utilizada análise preditiva são os modelos autorregressivos ou simplesmente **AR**. A definição formal de um modelo AR por [Brockwell e Davis 2002] é

**Definição 2.5.** *Uma série temporal  $\{X_t\}$  é um modelo  $AR(p)$  se é estacionária e satisfaz a seguinte equação para todo  $t$*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t \quad (2.26)$$

onde  $\{Z_t\} \sim WN(0, \sigma^2)$ .

Em condições gerais, pode-se mostrar que a solução da equação acima é um processo linear conforme definido na Equação 2.25. Por exemplo, para um AR(1) com  $|\phi| < 1$ ,  $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$  é a única solução existente.

Assim como o modelo de Markov oculto, utilizaremos os modelos autorregressivos como modelo preditivo no algoritmo de previsão adaptativo. No Capítulo 4, comparamos a acurácia da previsão do algoritmo adaptativo com modelos de Markov oculto e com modelos autorregressivos.

### 2.3.1 Estimação de parâmetros

Nos modelos autorregressivos também precisamos estimar seus parâmetros para descrever o processo. Formalmente, dado um conjunto de observações  $\mathbf{x} = x_1, \dots, x_n$ , queremos estimar  $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_p)$  e  $\sigma^2$  de forma que a função de autocovariância do modelo ( $\gamma_X(\cdot)$ ) se aproxime à das observações, ou seja, de forma que as suas características de segunda ordem fiquem próximas às das observações.

Para resolver este problema, apresentaremos o método de estimação de Yule-Walker. A idéia deste método é utilizar a autocovariância amostral  $\hat{\gamma}(h)$  para determinar os parâmetros  $\boldsymbol{\phi}$  e  $\sigma^2$ . Da Definição 2.5 temos que

$$\mathbf{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p \quad (2.27a)$$

$$\sigma^2 = \gamma_X(0) - \boldsymbol{\phi} \boldsymbol{\gamma}_p \quad (2.27b)$$

onde

$$\mathbf{\Gamma}_p = [\gamma_X(i-j)]_{p \times p}, \quad (2.28a)$$

$$\boldsymbol{\gamma}_p = (\gamma_X(1), \dots, \gamma_X(p))'. \quad (2.28b)$$

Desta forma, basta utilizar a função de autocorrelação amostral,  $\hat{\gamma}_X(h)$ , nas Equações 2.27 para estimar os parâmetros do modelo autorregressivo. Este conjunto de equações é freqüentemente chamado de estimadores de Yule-Walker.

A complexidade de cada evento de estimação é dada pelo cômputo da autocovariância amostral  $\hat{\gamma}(h)$  para  $0 \leq h \leq p$ , exigindo  $O(np - p^2)$  operações; mais a inversão da matriz de covariância  $\mathbf{\Gamma}_p$  com  $O(p^3)$  operações; e o produto na equação 2.27a com  $O(p^2)$  operações. Totalizando uma complexidade de  $O(p(n + p^2))$ .

## 2.4 Problema de previsão

Considerando um processo estocástico qualquer  $\{X_t\}$ . Formalmente, o problema de previsão é calcular o valor de  $X_{t+f}$ ,  $f \geq 0$ , dada as observações  $\{X_{t-1}, \dots, X_{t-p}\}$ . Podemos abordar este problema como um problema otimização, ou seja, realizar a previsão com o menor “erro” possível.

Assim, queremos uma função  $g : X_{t-1}, \dots, X_{t-p} \mapsto g(X_{t-1}, \dots, X_{t-p}) = \hat{X}_{t+f}$  tal que o **erro médio quadrático**,  $\text{MSE}^\dagger = E[(X_{t+f} - \hat{X}_{t+f})^2]$ , seja minimizado. A função  $g$  com tal propriedade é  $E[X_{t+f} | X_{t-1}, \dots, X_{t-p}]$  (ver Apêndice A.1).

Tome como exemplo um processo estacionário Gaussiano  $\{X_t\}$  com média  $\mu$  e variância  $\sigma^2$ . Supondo que queremos prever  $X_{t+f}$  com apenas a última observação  $X_{t-1}$ . Do Apêndice A.5, temos que

$$g(X_{t-1}) = E[X_{t+f} | X_{t-1}] = \mu + \rho(f+1)(X_{t-1} - \mu)$$

com erro médio quadrático  $E[(X_{t+f} - g(X_{t-1}))^2] = \sigma^2(1 - \rho(f+1)^2)$ . Assim, a medida que  $|\rho(f)| \rightarrow 1$ , o erro médio quadrático tende a zero e o preditor se reduz a replicação da observação  $X_{t-1}$ .

### 2.4.1 Preditores lineares

Entretanto, em um processo geral, obter  $E[X_{t+f} | X_{t-1}, \dots, X_{t-p}]$  nem sempre é possível ou imediato. Uma alternativa é escrever  $g$  como uma função linear, criando

---

<sup>†</sup>Acrônimo de *Mean Square Error*.



um preditor linear. Assim, nosso objetivo é calcular os coeficientes de tal forma a minimizar seu erro médio quadrático, obtendo o preditor linear ótimo:

$$\arg \min_{\phi_0, \dots, \phi_p} E[(X_{t+f} - \phi_0 - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})^2]. \quad (2.29)$$

onde  $p$  é denotado por ordem do preditor.

O ponto de mínimo pode ser obtido anulando as  $p + 1$  derivadas parciais em relação à  $\phi_0, \dots, \phi_p$ . Com isto teremos,

$$\mathbf{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p(f), \quad (2.30a)$$

$$\phi_0 = \mu \left( 1 - \sum_{i=1}^t \phi_i \right) \quad (2.30b)$$

onde

$$\boldsymbol{\gamma}_p(f) = (\gamma_X(f), \gamma_X(f+1), \dots, \gamma_X(f+(p-1)))'.$$

O erro médio da previsão  $E[X_{t+f} - \hat{X}_{t+f}]$  é nulo pela Equação 2.30(a). Enquanto o erro médio quadrático  $E[(X_{t+f} - P_t X_{t+f})^2]$  é dado por

$$E[(X_{t+f} - \hat{X}_{t+f})^2] = \gamma_X(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p(f). \quad (2.31)$$

É importante ressaltar que o modelo autorregressivo estimado para descrever a autocovariância do processo na Seção 2.3.1 é equivalente ao preditor linear ótimo para  $f = 1$  (Equações 2.27 e 2.30). Assim, utilizaremos os termos “preditores lineares” e “modelos autorregressivos” indistingüivelmente.

Na maioria dos casos, o erro médio quadrático de um preditor linear será maior do que a estimativa de previsão  $f(X_1, \dots, X_t) = E[X_{t+f} | (X_1, \dots, X_t)]$ . Todavia, este primeiro requer apenas o conhecimento da função de autocorrelação e da média do processo enquanto o segundo exige uma descrição mais específica do processo analisado, em termos da distribuição probabilística conjunta.

### 2.4.2 Análise de erro de dois preditores simples

Considerando uma previsão para o próximo instante de tempo, i.e.,  $f = 1$ . Uma estimativa simples da previsão é a média amostral (Apêndice A.3), dizemos que este é o **preditor média**. O erro médio quadrático para o preditor média é dado por  $E[(X_t - \bar{\mu}_X)^2] = \sigma_X^2 = \gamma_X(0)$ . Ou seja, o erro deste preditor é a variância do processo.

Outra estimativa simples para a previsão é a repetição da última observação do processo. Este preditor é denominado preditor replicador e seu erro médio quadrático é  $E[(X_t - X_{t-1})^2] = 2(\gamma_X(0) - \gamma_X(1))$ . Assim, conforme duas observações consecutivas se tornam mais correlacionadas, menor será o erro do deste preditor.

Caso o preditor tenha o conhecimento da autocorrelação do processo  $\{X_t\}$  e considerando um preditor linear de ordem 1 (AR(1)) por simplicidade, o erro médio quadrático é de  $E[(X_t - X_{t-1})^2] = \gamma_X(0) - \gamma_X(1)^2/\gamma_X(0)$ .

A Figura 2.3 mostra a relação dos erros destes 2 preditores simples em função da autocovariância  $\gamma_X(1)$  comparativamente ao preditor linear de ordem 1.

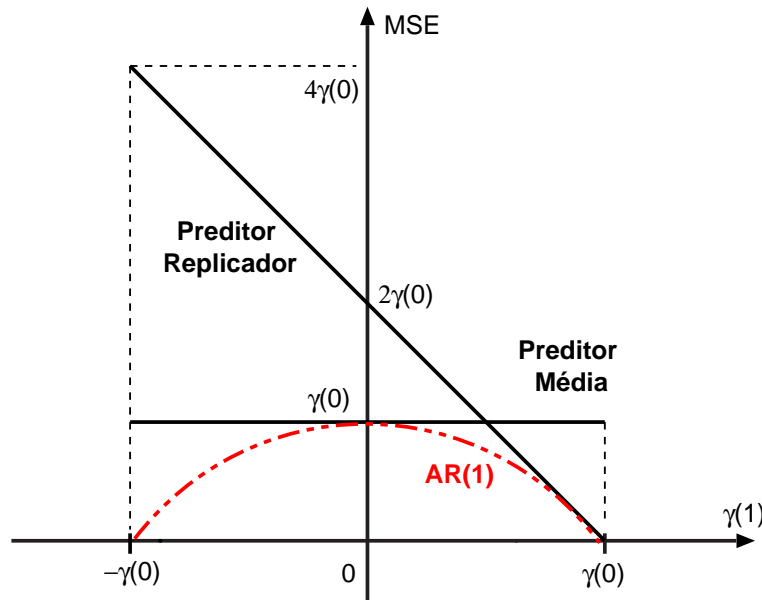


Figura 2.3: Relação entre os erros de 3 preditores simples

Naturalmente, a medida que dois valores consecutivos se tornam muito correlacionados ( $\gamma_X(1)/\gamma_X(0) \rightarrow 1$ ), o erro do replicador tende à zero. Enquanto dois valores consecutivos se tornam menos correlacionadas ( $\gamma_X(1)/\gamma_X(0) \rightarrow 0$ ), o erro do modelo autorregressivo de ordem 1 tende ao erro do preditor média porque não existirá relação entre estes valores ( $E[X_{t+1}|X_t] = E[X_{t+1}]$ ). Neste caso, o erro do replicador é ainda duas vezes pior.

## Capítulo 3

# Trabalhos relacionados e o algoritmo adaptativo de previsão

Definidos os conceitos teóricos no Capítulo 2, começamos este capítulo apresentando os trabalhos relacionados a esta dissertação. Em seguida, descreveremos de maneira detalhada o algoritmo adaptativo para previsão da taxa de perdas de pacotes mencionado no Capítulo 1.

### 3.1 Trabalhos relacionados

A variabilidade espacial e temporal das características da rede [Paxson 1997], tais como a taxa de perda de pacotes e o tempo gasto nas filas dos roteadores intermediários, motivou a aplicação de medições em protocolos adaptativos. Por exemplo, em transmissão de voz sobre IP, existem propostas na literatura de algoritmos que utilizam medições da taxa de perdas de pacotes para decidir como enviar informação redundante de forma mitigar os efeitos das perdas de pacotes [Boutremans e Boudec 2003, Bolot et al. 1999, Perkins et al. 1998]. No âmbito de controle de congestionamento, citamos os protocolos que realizam medições da taxa de perda e do retardo para controlar a taxa de transmissão [Mascolo et al. 2001,

Floyd et al. 2000, Brakmo e Peterson 1995].

Restringindo nosso contexto à perda de pacotes, o trabalho de [Yajnik et al. 1999] mostra que, em suas medições, o processo de perdas de pacotes apresenta autocorrelação em largas escalas de tempo. Como a autocorrelação de um processo está associada com sua previsibilidade, existe a possibilidade de protocolos adaptativos que tomem decisões de controle baseadas tanto em medições quanto em previsões da taxa de perda. Em [Silveira et al. 2006], o autor mostra que é possível aumentar a qualidade de serviço de uma aplicação VoIP através destas previsões.

Em [Su et al. 2004], os autores propõem um algoritmo para estimar a distribuição da taxa de perda futura **condicionada** às observações das perdas de pacotes recentes. Esta distribuição é obtida caracterizando o processo de perdas com uma cadeia de Markov de dois estados (modelo de Gilbert-Elliot simplificado definido no Apêndice A.2). Entretanto, este modelo é incapaz de representar as correlações temporais de larga escala do processo de perdas.

Já em [Salamatian e Vaton 2001], os autores modelam o processo de perdas através de um modelo de Markov oculto (**HMM**). Neste mesmo trabalho, os autores mostram que com poucos estados, em geral 4, o modelo HMM é capaz de representar as autocorrelações significativas do processo de perdas  $\{X_t\}$  observadas em [Yajnik et al. 1999].

Enquanto nos trabalhos [Duarte et al. 2003, Duarte 2003], os autores propõem um modelo de Markov oculto onde cada observação é o número agregado de perdas em um bloco de  $S$  pacotes. Além disto, naquele trabalho, o autor considera uma matriz de transição **A** particular motivado por suas medições realizadas à época. Apesar do número de parâmetros não ser grande devido à estrutura particular, o modelo proposto não é capaz de prever estatísticas do processo de perdas individuais além da taxa de perdas, tais como o tamanho médio da rajada de perdas por exemplo.

Nos trabalhos de [Silveira e de Souza e Silva 2006, Silveira et al. 2006], os auto-

res apresentam um modelo de Markov oculto hierárquico. Este modelo descreve as estruturas do processo de perdas, contidas em uma escala de tempo pequena, com um processo estocástico simples. Enquanto, sob escalas de tempo maiores, a dinâmica do processo de perdas é descrita pelas transições entre estes processos simples como ilustra a Figura 3.1(a). Em outras palavras,  $S$  observações de perda são geradas a partir de uma cadeia de Markov de dois estados e, após este lote de observações, haverá uma transição para outra cadeia de Markov. A Figura 3.1(b) exemplifica a estrutura deste modelo com 3 estados ocultos.

## 3.2 Medida da previsão

Daqui em diante, consideraremos que o processo de perda  $\{X_t\}$  indica a observação de perda do  $t$ -ésimo pacote transmitido pela aplicação. Assumindo valor 1 (0) caso o pacote transmitido seja perdido (entregue).

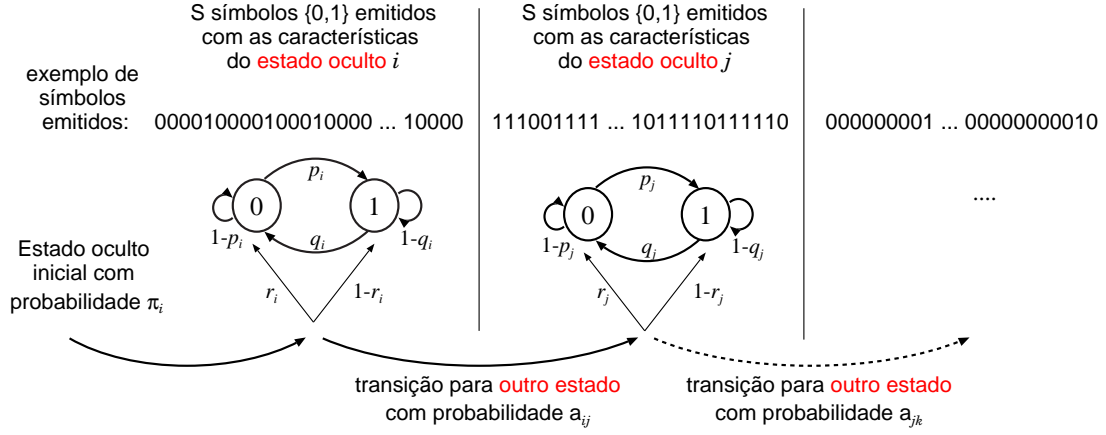
Sobre o processo de perdas, definimos o processo de taxa de perdas  $\{R_t^F\}$ . Onde seus valores representam a fração de pacotes perdidos nas  $t, \dots, (t + F - 1)$  transmissões, i.e.

$$R_t^F = \frac{1}{F} \sum_{f=0}^{F-1} X_{t+f}. \quad (3.1)$$

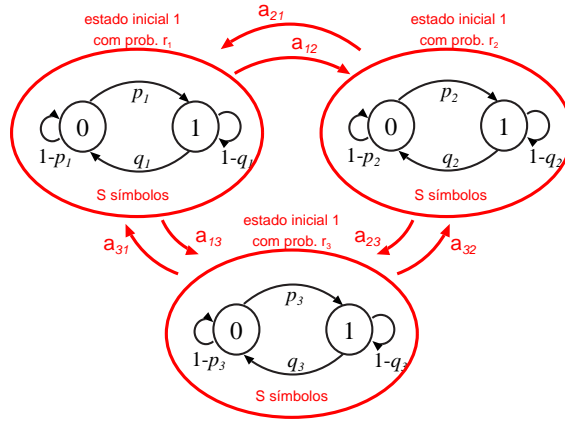
Logo,  $R_t^F$  pode assumir  $F + 1$  valores  $(0, 1/F, \dots, (F - 1)/F, 1)$ . Um exemplo de um caminho amostral hipotético do processo de perdas e da respectiva série de taxa de perdas para  $F = 4$  podem ser visto na Figura 3.2.

De acordo com o Capítulo 1, iremos prever os valores da taxa de perdas em uma janela de tempo futura, ou equivalentemente, **janela de previsão**. Assim, o problema de previsão será estimar os valores de  $R_t^F$ , condicionados às observações recentes de perdas  $\{X_t\}$ . A previsão de  $R_t^F$  será denotada por  $\hat{R}_t^F$  e a série da taxa de perdas previstas por  $\{\hat{R}_t^F\}$ .

A Figura 3.3 mostra uma janela de previsão hipotética. Neste exemplo, temos em uma janela de previsão três estimativas da taxa de perda real. Particularmente



(a) Dinâmica do modelo HMM considerado



(b) Exemplo da estrutura com 3 estados do modelo HMM hierárquico

Figura 3.1: Modelo HMM hierárquico considerado nesta dissertação

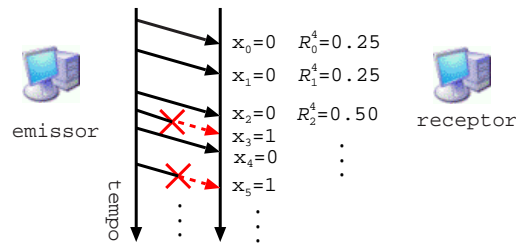


Figura 3.2: Taxa de perda  $R_t^F$

em nossa análise preditiva, para cada janela de previsão há apenas uma estimativa da taxa de perda.

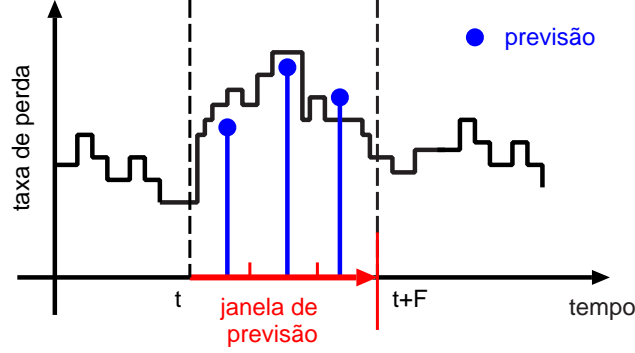


Figura 3.3: Janela de previsão  $F$  hipotética

Com isto, a janela de previsão é um parâmetro crítico devido à sua influência no desempenho da aplicação. Por exemplo, um mecanismo de correção de erros (**FEC**) pode ter sua previsão muito alterada se fizermos a previsão da taxa considerando-a constante na janela de previsão, e portanto sem considerar variações desta taxa em um tempo relativamente curto.

A Figura 3.4 mostra este aumento da variabilidade da taxa de perdas para janelas de previsões menores em um processo de perdas qualquer. Por outro lado, em janelas de previsões maiores, a variabilidade diminui porque a taxa de perda converge para  $\mu_X = E[X_t]$  pela lei dos grandes números [Ross 1982].

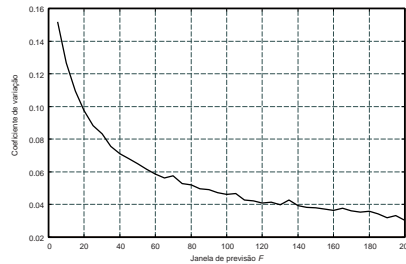


Figura 3.4: Coeficiente de variação da taxa de perda para diferentes valores de  $F$



### 3.3 Algoritmo adaptativo para previsão da taxa de perda

Para obter as previsões da taxa de perda, utilizaremos o algoritmo adaptativo de previsão\* proposto no trabalho de [Silveira e de Souza e Silva 2006]. Este algoritmo é responsável tanto por estimar quanto atualizar o modelo preditivo às variações do processo de perda ao longo do tempo e, em cada previsão, obter uma estimativa da taxa de perdas futura condicionada às observações das perdas passadas.

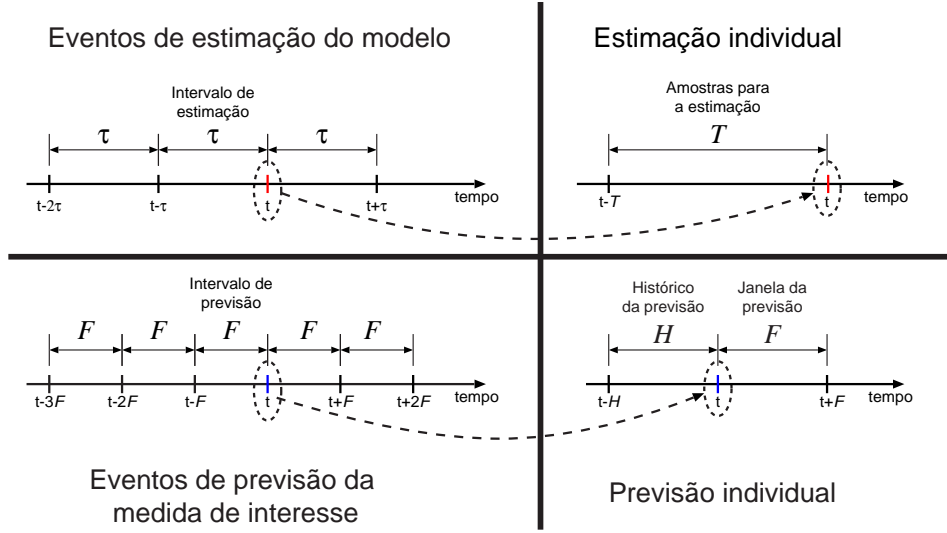


Figura 3.5: Algoritmo adaptativo em função dos parâmetros de estimação ( $T, \tau$ ) e de previsão ( $H, F$ )

A Figura 3.5 ilustra como estes eventos ocorrem individualmente ao longo do tempo: a cada  $\tau$  observações do processo de perdas  $\{X_t\}$ , o modelo preditivo é atualizado às mudanças que possam ocorrer no processo de perdas. Caso o processo seja estacionário, um único modelo preditivo corretamente estimado pode ser utilizado ( $\tau \rightarrow \infty$ ). Na prática, o que se observa são intervalos de tempo onde o processo  $\{X_t\}$  apresenta características estacionárias (e.g. [Ji et al. 2004]). Paralelamente, a cada  $F$  observações do processo de perdas, uma previsão da taxa de perdas,  $R_t^F$ ,

\*Usaremos os termos algoritmo adaptativo de previsão, ou apenas algoritmo de previsão, e mecanismo de previsão adaptativa indistingüivelmente.

condicionada às últimas  $H$  observações é realizada.

Nesta dissertação consideramos dois modelos preditivos: os modelos de Markov oculto e os modelos autorregressivos. A seguir, apresentamos como o algoritmo adaptativo está associado a estes modelos.

### 3.3.1 Algoritmo adaptativo com modelos de Markov ocultos

#### Estimação

Em cada evento de reestimação dos parâmetros  $\lambda$  do modelo HMM, utiliza-se o algoritmo de Baum-Welch descrito na Seção 2.2.1, considerando as  $T$  últimas amostras do processo de perdas.

#### Previsão

Para obter a previsão da taxa de perdas  $R_t^F$ , faremos uma análise transiente ([de Souza e Silva e Gail 2000]) do modelo HMM com estado inicial condicionado aos eventos de perdas  $X_{t-H}, \dots, X_{t-1}$ . Assim teremos uma estimativa da distribuição probabilística de  $R_t^F$ . Isto é, para calcular a distribuição de  $R_t^F$  condicionada às últimas observações, temos a seguinte equação

$$\begin{aligned} P(R_t^F = j | X_{t-H}, \dots, X_{t-1}) &= \sum_{\forall y_{t-1}} P(R_t^F = j, Y_{t-1} | X_{t-H}, \dots, X_{t-1}) \\ &= \sum_{\forall y_{t-1}} P(R_t^F = j | Y_{t-1}, X_{t-H}, \dots, X_{t-1}) P(Y_{t-1} | X_{t-H}, \dots, X_{t-1}) \\ &= \sum_{\forall y_{t-1}} P(R_t^F = j | Y_{t-1}) P(Y_{t-1} | X_{t-H}, \dots, X_{t-1}). \end{aligned} \quad (3.2)$$

A última igualdade formaliza o algoritmo de previsão utilizando um modelo de Markov oculto em duas etapas: **1**– obter a distribuição do estado oculto inicial **condicionado** às observações  $X_{t-H}, \dots, X_{t-1}$  através do procedimento *forward*; **2**– estimar a distribuição de  $R_t^F$  dado o estado oculto  $Y_{t-1}$ . Como  $R_t^F$  é uma função dos símbolos emitidos pelo HMM, a segunda etapa consiste em realizar análise transiente no HMM para as transições  $t, \dots, t + F - 1$ . Em [Silveira 2006], o autor apresenta um algoritmo recursivo para calcular esta distribuição da segunda etapa utilizando apenas operações matriciais.

De posse da estimativa de  $P(R_t^F = j | X_{t-H}, \dots, X_{t-1})$ , utilizaremos como previsão  $\hat{R}_t^F = E[R_t^F | X_{t-H}, \dots, X_{t-1}]$ . Entretanto, qualquer estatística poderia ser utilizada (e.g.  $\arg \max_j P(R_t^F = j | X_{t-H}, \dots, X_{t-1})$ ). A motivação desta escolha é minimizar o erro médio quadrático (Apêndice A.1).

### 3.3.2 Algoritmo adaptativo com modelos autorregressivos

Com o modelo autorregressivo, o algoritmo adaptativo segue o esquema ilustrado na Figura 3.5 com pequenas alterações. Por uma limitação do modelo, não podemos utilizar a série binária de perdas  $\{X_t\}$ , o que nos leva a considerar apenas a série  $\{R_t^F\}$ .

#### Estimação

O algoritmo adaptativo com este modelo preditivo é similar: a cada  $\tau'$  observações de  $\{R_t^F\}$ , ou equivalentemente  $\tau' \cdot F$  amostras de  $\{X_t\}$ , os parâmetros  $\phi$  são reestimados considerando as  $T'$  últimas amostras de  $\{R_t^F\}$ ,  $T' \cdot F$  amostras de  $\{X_t\}$ . Em cada evento de estimação, o algoritmo descrito na Seção 2.3.1 é utilizado.

#### Previsão

Em relação ao cálculo da previsão  $R_{t+1}^F$ , considera-se um histórico de  $H = p$  amostras de  $\{R_t^F\}$  ( $p \cdot F$  amostras de  $\{X_t\}$ ). Sendo  $p$  a ordem de autorregressão do modelo autorregressivo. A Figura 3.6 mostra um evento de previsão hipotético quando o histórico de previsão de ambos os modelos são equivalentes.

Apenas para facilitar a leitura, reescreveremos a Equação 2.26 do processo AR em termos da série  $\{R_t^F\}$  como

$$R_t^F - \phi_1 R_{t-F}^F - \phi_1 R_{t-2F}^F - \dots - \phi_p R_{t-pF}^F = Z_t, \quad (3.3)$$

onde,  $\{Z_t\} \sim WN(0, \sigma^2)$ .

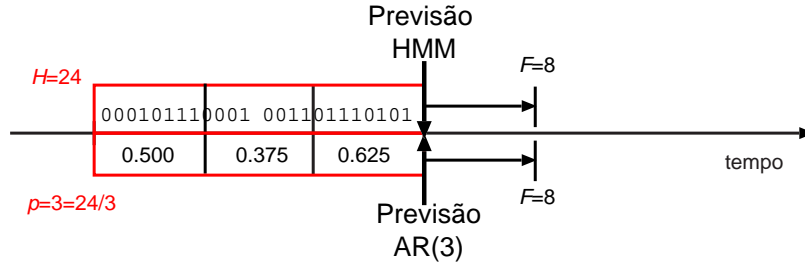


Figura 3.6: Relação entre um evento de previsão hipotético com modelo HMM e modelo AR(3)

Como o objetivo é realizar a previsão do instante de tempo seguinte, denotemos

$$\hat{R}_t^F = \begin{cases} 0, & \text{se } t < p, \\ \hat{R}_t^F, & \text{se } t \geq p \end{cases},$$

onde  $\hat{R}_t^F = \phi_1 R_{t-F}^F + \dots + \phi_p R_{t-pF}^F$  é o preditor linear de ordem  $p$  visto na Seção 2.4.1.

Em termos computacionais, tanto a estimação quanto o cálculo da previsão no algoritmo adaptativo considerando os modelos de Markov oculto e autorregressivo são realizados rapidamente nos computadores atuais. Assim, é viável a sua utilização em aplicações de tempo real, como a implementação na aplicação de Voz sobre IP *FreeMeeting* [FreeMeeting 2007] realizada em [Prezzi e Murai 2007] por exemplo. Portanto, não enfatizaremos o custo computacional de ambos os modelos.

# Capítulo 4

## Resultados experimentais

Neste capítulo apresentamos os principais experimentos realizados com o algoritmo adaptativo descrito na Seção 3.3. De acordo com o Capítulo 1, os objetivos gerais destes experimentos são: **1**– avaliar a capacidade de previsão deste algoritmo adaptativo utilizando os modelos de Markov ocultos, modelos autorregressivos, e alguns preditores simples; **2**– verificar a eficácia do algoritmo de previsão em relação aos seus parâmetros: **i**) histórico de estimação  $T$ , **ii**) intervalo de reestimação  $\tau$ , **iii**) histórico de previsão  $H$  e **iv**) janela de previsão  $F$ , com os modelos e cenários aqui considerados; **3**– utilizar cenários de experimentação mais diversificados dos aqueles previamente considerados.

Os experimentos estão organizados da seguinte forma: inicialmente, o processo de perdas utilizado como entrada para o algoritmo adaptativo será um modelo analítico de forma a facilitar a análise dos resultados. Em seguida, utilizaremos um conjunto de medições de perdas de pacotes observadas por uma aplicação VoIP com codificação **PCM** e sem detecção de silêncio implementado em [FreeMeeting 2007] na *Internet*. Por fim, analisaremos processos de perdas obtidos em cenários simulados onde podemos controlar os experimentos.

## 4.1 Capacidade de previsão do algoritmo adaptativo

Para avaliar a capacidade de previsão e verificar a eficácia do algoritmo de previsão em relação aos seus parâmetros, definimos um conjunto de medidas quantitativas a seguir. Apesar de utilizarmos um conjunto pequeno de medidas, trabalhamos com inúmeras outras medidas e selecionamos aquelas que se mostraram mais adequadas.

### Primeiro conjunto de medidas

1. A primeira medida utilizada é o, já mencionado, erro médio quadrático  $\text{MSE} = E[(\hat{R}_t^F - R_t^F)^2]$ . Do Apêndice A.1, o erro médio quadrático está limitado inferiormente pelo erro do preditor ótimo

$$\text{MSE}^* = E[(R_t^F)^2] - \sum_{X_{t-H}, \dots, X_{t-1}} E[(R_t^F) | X_{t-H}, \dots, X_{t-1}]^2 P[X_{t-H}, \dots, X_{t-1}].$$

Entretanto, em situações reais, este limite não é sabido porque a distribuição probabilística do processo  $\{X_t\}$  dificilmente é conhecida.

2. Como segunda medida de desempenho, temos a taxa de acerto relativa. Ela será muito útil por ter uma interpretação imediata. Além disto, em situações práticas, para uma aplicação um erro arbitrado pode ser aceitável.

Formalmente, seja a variável aleatória que indica o acerto relativo

$$I_t^\alpha = \begin{cases} 1 & \hat{R}_{t+1} \in [R_{t+1}(1 - \alpha), R_{t+1}(1 + \alpha)], \\ 0 & \text{caso contrário.} \end{cases}$$

A taxa de acerto relativa é definida como  $E[I_t^\alpha] = \mathbf{I}(\alpha)$ .

Como  $R_t^F$  assume  $F + 1$  valores, um “preditor aleatório equiprovável” tem uma taxa de acerto de  $I_m(0) = 1/(F + 1)$ .

3. A ultima medida de desempenho é a correlação cruzada, denotada por  $\mathbf{COR} = \hat{\rho}_{R_t^F \hat{R}_t^F}$  (Equação 2.5), entre a série  $R_t^F$  e  $\hat{R}_t^F$ . Com ela, temos a noção se a

variabilidade de  $\hat{R}_t^F$  está associada com a variabilidade de  $R_t^F$ . Com ela, temos a noção se a variabilidade de  $\hat{R}_t^F$  está associada com a variabilidade de  $R_t^F$ .

## Segundo conjunto de medidas: selecionar trechos com variação

Conforme discutido no Capítulo 1, o desempenho da aplicação aumenta com a sua capacidade de prever variações da taxa perda. Entretanto, se olharmos o caso onde a taxa de perda é nula durante um longo período de tempo com elevações bruscas de curta duração (picos) como exemplificado na Figura 4.1, um preditor  $\hat{R}_t^F \equiv 0$  seria muito bom de acordo com as medidas acima mas seria inútil para a aplicação.

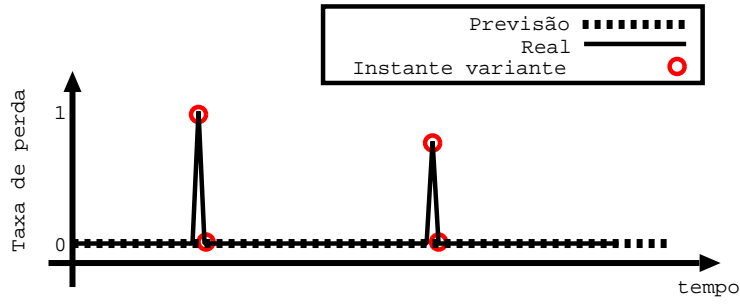


Figura 4.1: Preditor constantemente nulo em um processo com picos de perdas

Evitaremos estas situações particionando o tempo de observação em instantes que apresentam ou não uma variabilidade arbitrada. Um instante de tempo  $t$  será dito variante quando

$$R_t^F \notin \bigcup_{i=1..k} [R_{t-i}^F - \delta, R_{t-i}^F + \delta]$$

ou invariante caso contrário.

Considerando apenas a previsão e a taxa de perda real nos instantes variantes, definimos o mesmo conjunto de medidas definidas anteriormente. Isto é, obtemos as três medidas de desempenho condicionadas à variabilidade da taxa de perda.

## 4.2 Erro de estimação do modelo HMM

Avaliaremos a estimação dos parâmetros do modelo preditivo HMM através da sua capacidade em representar dois descritores de primeira ordem do processo de perdas. Sendo o primeiro um descritor de média e o segundo um descritor da estrutura temporal das perdas individuais – uma vez que utilizamos o modelo HMM para realizar previsão de perdas, é importante que ele descreva a estrutura do processo de perdas.

Consideremos duas estatísticas auxiliares no processo de perdas  $\{X_t\}$ . A fração de perda amostral  $l_k(t)$

$$l_k(t) = \frac{\sum_{i=1}^k X(t-i)}{k} \quad (4.1)$$

que, de fato, é o estimador da média  $\mu_X$  do processo  $\{X_t\}$ . E a rajada média amostral  $r_k(t)$ , definida como

$$r_k(t) = \frac{\sum_{i=1}^{n(k)} Z(t-i)}{n(k)}, \quad (4.2)$$

onde  $\{Z_t\}$  é a série que representa o tamanho do segmento de perdas consecutivas, ou apenas **rajada de perdas**, construída a partir de  $\{X_t\}$  (Figura 4.2) e  $n(k)$  é o número de rajadas nas últimas  $k$  amostras do processo de perdas.

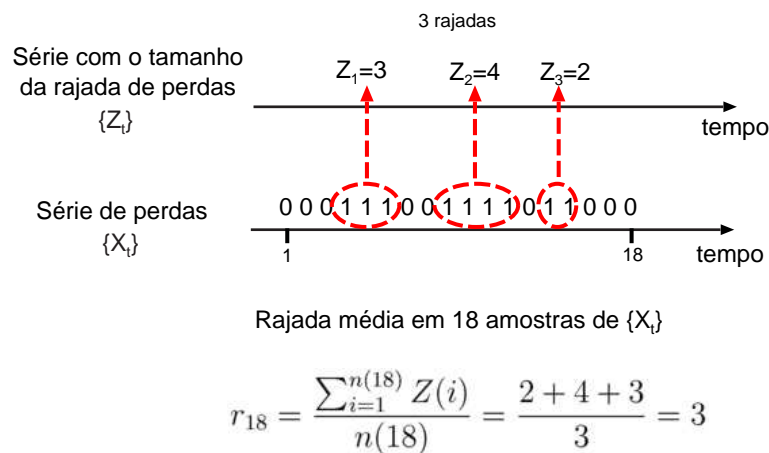


Figura 4.2: Exemplo de uma rajada média amostral

Para avaliar o processo de estimação dos parâmetros do HMM, verificamos a partir de que valor de  $T$ , as respectivas estatísticas estimadas convergem para 4.1



e 4.2 – tendo  $k = T$ . Assim, poderemos quantificar o erro da estimação do HMM através do erro dos estimadores amostrais e obter um “intervalo de confiança” para a estimação.

Note que a fração de perda estimada pelo HMM com  $T$  amostras do processo  $\{X_t\}$  é calculada a partir da seguinte equação

$$\hat{\mu}_X^T = \pi \hat{\mu}' \quad (4.3)$$

onde  $\pi$  é o vetor linha com a distribuição estacionária dos estados ocultos e  $\hat{\mu}'$  é um vetor coluna no qual o  $i$ -ésimo elemento representa a fração de perda no estado oculto  $i$ , vide Equação A.3. E a rajada média estimada pelo HMM é calculada a partir de

$$\hat{\mu}_Z^T = \pi \hat{\mu}'_Z \quad (4.4)$$

onde  $\hat{\mu}'_Z$  é um vetor coluna no qual o  $i$ -ésimo elemento representa a rajada média condicionada ao estado oculto  $i$  (Equação A.5).

## 4.3 Resultados usando modelos analíticos

Começaremos avaliando a sensibilidade dos parâmetros do algoritmo adaptativo e verificando a sua capacidade de previsão quando consideramos que o processo de perdas será dado por modelo analítico. O objetivo é obter resultados analíticos que não seriam possíveis com medições de um processo de perda.

### Modelo de Gilbert-Elliot

Os resultados deste experimento foram obtidos utilizando o modelo de Gilbert-Elliot simplificado visto no Apêndice A.2. Arbitramos um Gilbert-Elliot simplificado com  $p = 0.005$  e  $q = 0.05$ . Estes valores foram escolhidos de forma a fixar a fração de perda em 10% e rajadas de perdas média de 20 pacotes. A Figura 4.3 mostra a função de autocorrelação da taxa de perdas deste processo.

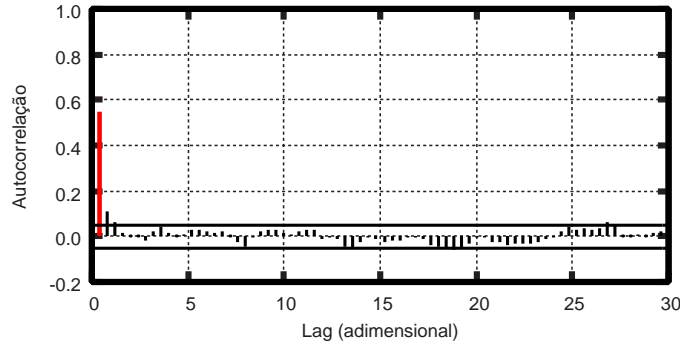


Figura 4.3: Autocorrelação da taxa de perda ( $F = 20$ ) para um modelo de perdas de Gilbert-Elliot simplificado com a fração de perda de 10% e rajadas de perdas média de 20 pacotes

### Estimação do modelo HMM

Como o processo de perdas é estacionário neste caso, não é necessário reestimar-mos o modelo periodicamente, então o papel de  $\tau$  neste cenário é irrelevante. Os

resultados adiante foram obtidos com  $N = 2$  estados - número mínimo para um HMM.

A respeito da fração de perda estimada pelo HMM,  $\hat{\mu}_X$ , utilizaremos o teste de hipótese:

1. Hipótese nula: a distribuição da fração de perda estimada em  $T$  amostras de  $\{X_t\}$  pelo HMM,  $\hat{F}_{\hat{\mu}}^n(l)$ , converge para a distribuição da fração de perda amostral  $F_{l_T}(l)$ , onde  $\hat{F}_{\hat{\mu}}^n(l) = (\text{Número de valores de } \hat{\mu}_X^T \text{ menores do que } l)/n$ . No modelo de Gilbert-Elliot, a distribuição da média amostral é computada recursivamente.
2. Para validar a hipótese nula, utilizamos o teste de Kolmogorov-Smirnov, sendo a estatística de teste

$$D_n = \max_l |\hat{F}_{\hat{\mu}}^n(l) - F_{l_T}(l)|.$$

O teste de Kolmogorov-Smirnov tem grande valia porque a distribuição **exata** da estatística de teste é função apenas de  $n$  e independente da distribuição de  $F_{l_T}(l)$  enquanto outros testes fornecem apenas aproximações mesmo com um número  $n$  razoável de amostras [Trivedi 2001].

3. Rejeitaremos a hipótese nula se o valor da estatística de teste estiver acima do valor crítico  $d_{n;\alpha}$  para um nível de significância  $\alpha$ .

Experimentalmente, verificamos que a fração de perda do modelo de Markov oculto estimado converge em distribuição para a estimador amostral  $l_T$  quando  $T > 600$  amostras. Assim, teremos que o erro de estimação para a fração de perda será aproximadamente o erro do estimador amostral. A Figura 4.4 mostra o erro médio quadrático da estimação e o erro médio quadrático da fração de perda amostral  $l_T$ .

Como o processo de perdas não é independente, o estimador amostral da fração de perda tem o seguinte intervalo de confiança (ver Apêndice A.3)

$$P \left[ |\hat{\mu}_X^T - \mu_X| \leq 1,96 \frac{\sigma_X}{\sqrt{T/K}} \right] \geq 95\%$$

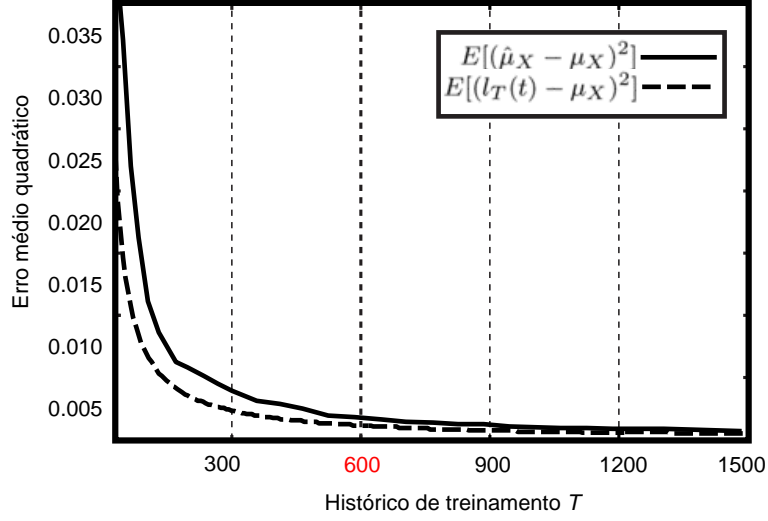


Figura 4.4: Comparação entre o erro médio quadrático da média amostral e o erro médio quadrático da média do modelo HMM estimado para  $p = 0.005$  e  $q = 0.05$

sendo  $K \approx 2$  para o processo de Gilbert-Elliot com  $p = 0.005$  e  $q = 0.05$  e 600 amostras de estimação (Equação A.8).

A respeito da rajada média estimada pelo HMM  $\hat{\mu}_Z^T$ , utilizaremos o teste de hipótese:

1. Hipótese nula: a distribuição da rajada média estimada em  $T$  amostras de  $\{X_t\}$  pelo HMM,  $\hat{F}_{\hat{\mu}_Z}^n(z)$ , converge para a distribuição da rajada média amostral  $F_{R_T}(z)$ .
2. Para validar a hipótese nula, utilizamos o teste de Kolmogorov-Smirnov descrito anteriormente.
3. Rejeitaremos a hipótese nula se o valor da estatística de teste estiver acima do valor crítico  $d_{n;\alpha}$  para um nível de significância  $\alpha$ .

Sobre a estimação da rajada média, o processo  $\{Z_t\}$  é independente e identicamente distribuído com distribuição Geométrica( $q$ ). Logo, a rajada média amostral  $R_T(t) \sim N\left(\mu_Z, \frac{\sigma_Z^2}{E[n(T)]}\right)$ . Neste experimento, a rajada média estimada no HMM

converge em distribuição para a rajada média amostral quando  $T > 700$ . O intervalo de confiança da rajada média amostral é

$$P \left[ |\hat{\mu}_Z^T - \mu_Z| \leq 1,96 \frac{\sigma_Z}{\sqrt{(T/(\mu_Z + \mu_W))}} \right] \geq 95\%.$$

## Previsão

No que se segue, verificamos a capacidade de previsão do algoritmo adaptativo com modelos HMM e AR comparativamente ao preditor ótimo (apresentado no Apêndice A.1). Como o processo de perdas é uma cadeia de Markov, o preditor ótimo  $E[R_t^F | X_{t-H}, \dots, X_{t-1}]$  se reduz à  $E[R_t^F | X_{t-1}]$ . Ou seja, apenas uma observação é relevante para a previsão. Assim, o erro médio quadrático pode ser derivado de  $E[(R_t^F - E[R_t^F | X_{t-1}])^2]$  à

$$\text{MSE} = E[(R_t^F)^2] - \sum_{x=0,1} (E[R_t^F | X_{t-1} = x])^2 P[X_{t-1} = x].$$

Ambos os termos podem ser calculados com métodos numéricos com os parâmetros da cadeia previamente escolhidos. A Figura 4.5 compara o MSE do algoritmo adaptativo associado ao modelo HMM e ao modelo AR com o preditor ótimo variando a janela de previsão  $F$ . O valor utilizado como histórico de previsão foi aquele que apresentou menor MSE.

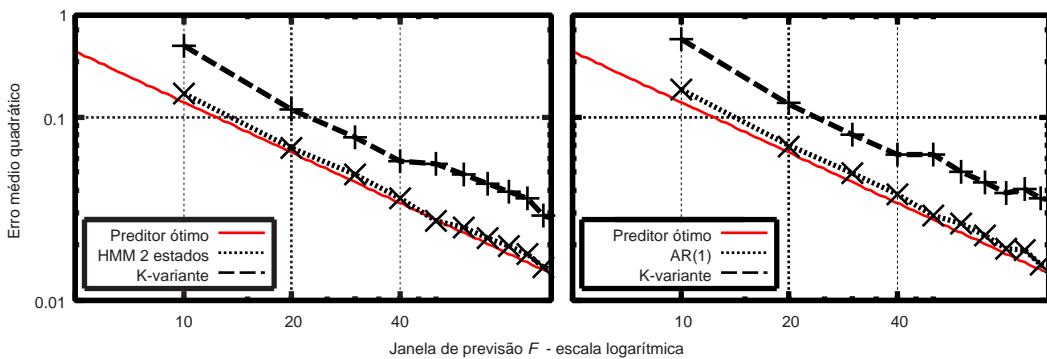


Figura 4.5: MSE dos modelos HMM, AR, média e replicador

Com janelas de previsão maiores o erro diminui porque a variabilidade da taxa de perda é inversamente proporcional à  $F$ , pela lei dos grandes números.

Para uma janela de previsão  $F = 20$ , a Figura 4.6 mostra que as medidas de desempenho do preditor não são influenciadas pelo valor do histórico de previsão  $H$ . Na Tabela 4.1, temos as medidas de desempenho dos preditores. Como veremos no próximo experimento, este tamanho da janela está relacionado à taxa de transmissão de uma aplicação VoIP simplificada e uma janela de previsão de 2  $RTT$ 's.

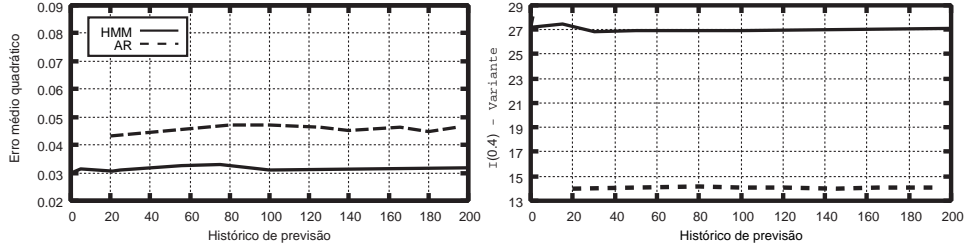


Figura 4.6: Relação entre a previsão e o histórico de previsão  $H$

Tabela 4.1: Comparação de medidas de desempenho para uma janela de previsão  $F = 20$ ,  $H = 20$  pacotes ( $p = 1$ ) e  $T = 1000$  pacotes

Modelo	MSE Variante	COR Variante	$I(0.4)$ Variante
HMM	0.13032	0.50945	28%
AR(1)	0.17771	0.17581	13%
Replicador	0.26301	0.05178	5%
Média	0.22495	0.09723	1%

Dos resultados apresentados, dois pontos devem ser destacados: Neste experimento, apenas 1 amostra é suficiente para que o HMM tenha um MSE igual ao do preditor ótimo.

Através dos segmentos de variabilidade destacados na Figura 4.7, vemos que o preditor HMM consegue prever melhor a taxa de perda. Enquanto o modelo AR atua apenas como um filtro linear na taxa de perdas, como evidencia a figura 4.8.

Outras observações experimentais não reportadas foram: o desempenho do modelo HMM com um número maior de estados ocultos foi similar ao modelo de dois estados apresentado. A influência do histórico de estimação  $T$  não foi notada no desempenho do preditor quando  $T > 500$ . Logo, o limite inferior obtido anteriormente

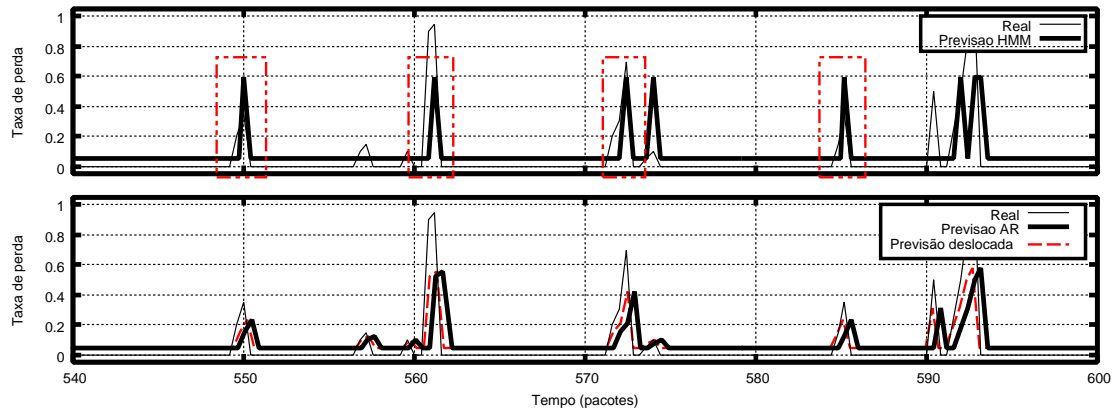


Figura 4.7: Previsão de um modelo de perdas de Gilbert-Elliot simplificado

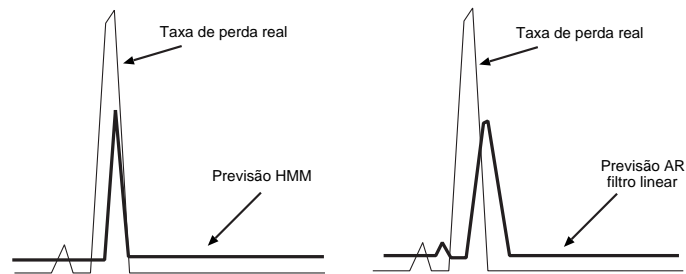


Figura 4.8: Previsão de um modelo de perdas de Gilbert-Elliot simplificado

era robusto.

Nossa avaliação será baseada no modelo Gilbert-Elliot. Apesar de ser pouco realista, este é um modelo analítico simples que exibe alguma estrutura temporal, além de ser muito utilizado.

## 4.4 Cenários com *traces*

Nesta segunda parte, avaliaremos a capacidade de previsão e a sensibilidade dos parâmetros do mecanismo de previsão adaptativa em um conjunto de medições reais obtidas na *Internet*.

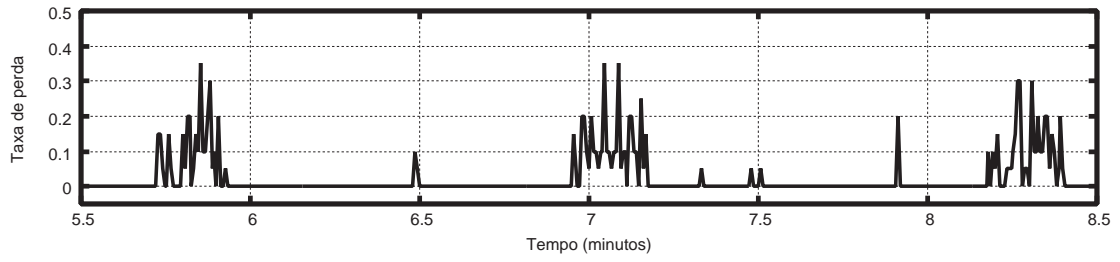
As medições foram obtidas com auxílio do módulo *Traffic Generator* da ferramenta Tangram-II [de Souza e Silva et al. 2006] e utilizadas nos trabalhos de [Duarte 2003, Duarte et al. 2003] e em [Silveira e de Souza e Silva 2006, Silveira et al. 2006]

A cada intervalo de tempo de 20 milissegundos, um pacote com 324 *bytes* de dados é enviado pelo gerador de tráfego para um destino arbitrado – somando os cabeçalhos do par UDP-IP (8+20 *bytes*) teremos uma taxa constante de 140.80 Kbps, desprezível se comparada com a capacidade dos canais envolvidos. No destinatário, são feitos os registros (*trace*) dos pacotes perdidos (1) e recebidos (0) compondo nosso processo de perdas  $\{X_t\}$ . A motivação era emular uma aplicação de voz sobre IP (VoIP) simplificada com codificação **PCM** na *Internet*. Como os pacotes estão sendo enviados a taxa constante, podemos estabelecer uma relação entre as observações e o tempo – uma a cada 20 milissegundos.

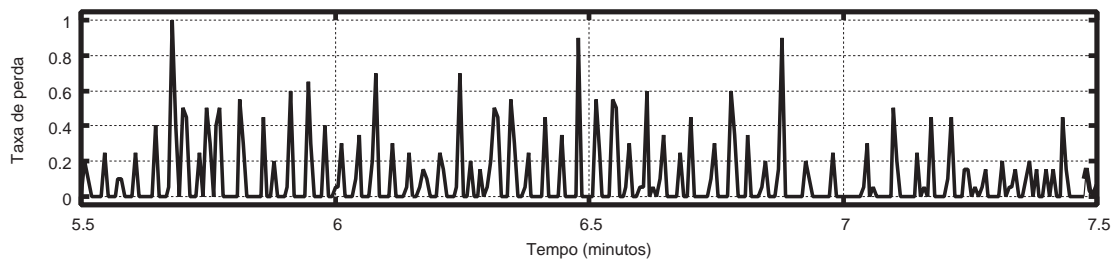
As medições ocorreram entre 4 instituições acadêmicas, sendo duas no Brasil (UFRJ e UFMG) e duas nos Estados Unidos (UMass e UMd), durante cerca de 1 hora em três horários de alta utilização ao longo do dia. Para analisar alguns *traces*, escolhemos aqueles onde havia alguma característica interessante na taxa de perda.

Em geral, observamos variações periódicas na taxa de perda (Fig. 4.9(a)), variabilidade intensa (Fig. 4.9(b)) e taxas de perdas constantes com altas variações de curto tempo (Fig. 4.9(c)). A Figura 4.10(a) quantifica a variabilidade da taxa de perda  $\{R_t^F\}$  através do coeficiente de variação em função da janela  $F$ . Enquanto as respectivas funções de autocorrelação amostral (Equação 2.10) do processo de perda, fundamentais para a previsão, mostram dependências temporais bem diferentes (Figura 4.10(b)).

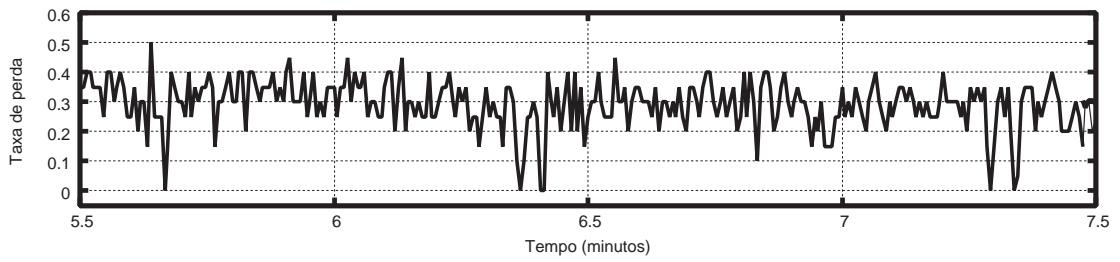




(a) Taxa de perda com rajadas periódicas em uma medição entre UMD e UFRJ

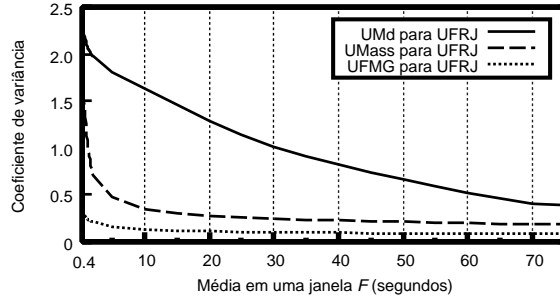


(b) Taxa de perda com alta variabilidade em uma medição entre UMass e UFRJ

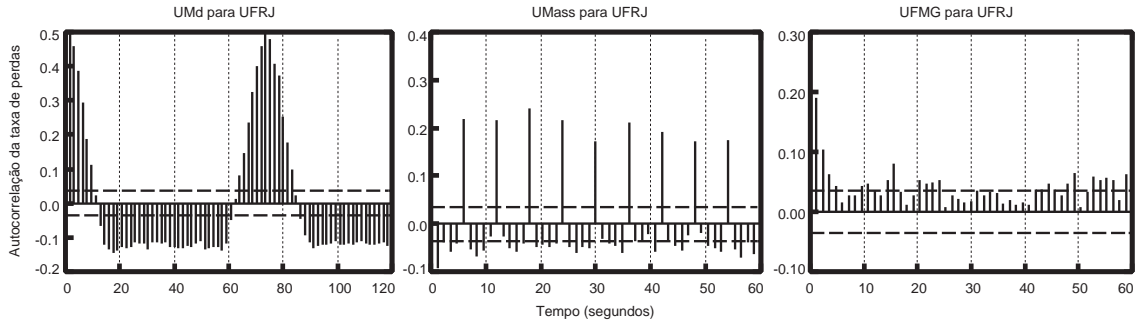


(c) Taxa de perda elevada em uma medição entre UFMG e UFRJ

Figura 4.9: Taxa de perda de três *traces* selecionados para uma janela de previsão  $F = 400$  milissegundos



(a) Coeficiente de variação da taxa de perda  
das Figuras 4.9(a-c)



(b) Autocorrelação do processo de perdas das Figuras 4.9(a-c)

Figura 4.10: Características da taxa de perda *traces* selecionados

Para obter os resultados do experimento, primeiro estimamos o número de estados ocultos seguindo a metodologia da Seção 2.2.2. Pelo resultado do experimento anterior, supomos que a partir  $T = 700$  amostras, a taxa de perda e a rajada média de perdas do modelo HMM convergem para distribuição amostral. Desta forma, observamos que os erros dos estimadores amostrais (Eq. 4.1 e 4.2) são aproximadamente constantes a partir de  $T = 6000$ ,  $T = 500$  e  $T = 3000$  amostras, respectivamente às Figuras 4.9(a-c).

Daqui em diante, iremos assumir que a janela de previsão é proporcional a  $2RTT's$  – um  $RTT$  de aproximadamente 200 milissegundos é razoável para pares entre o Brasil e EE.UU.. Faremos desta forma para que a aplicação tenha a previsão da taxa de perda para o próximo  $RTT$ . O motivo para prevermos o dobro

do desejado é porque, no instante de tempo  $t$ , a aplicação detém as observações do processo de perda somente até  $t - RTT$ , ou seja, atrasadas de 1  $RTT$ . Para realizar a previsão do tempo futuro, é necessário descontarmos este atraso. Logo, totalizam-se  $2RTT$ 's. A Figura 4.11 esquematiza esta relação entre a previsão e a informação do processo de perdas disponível.

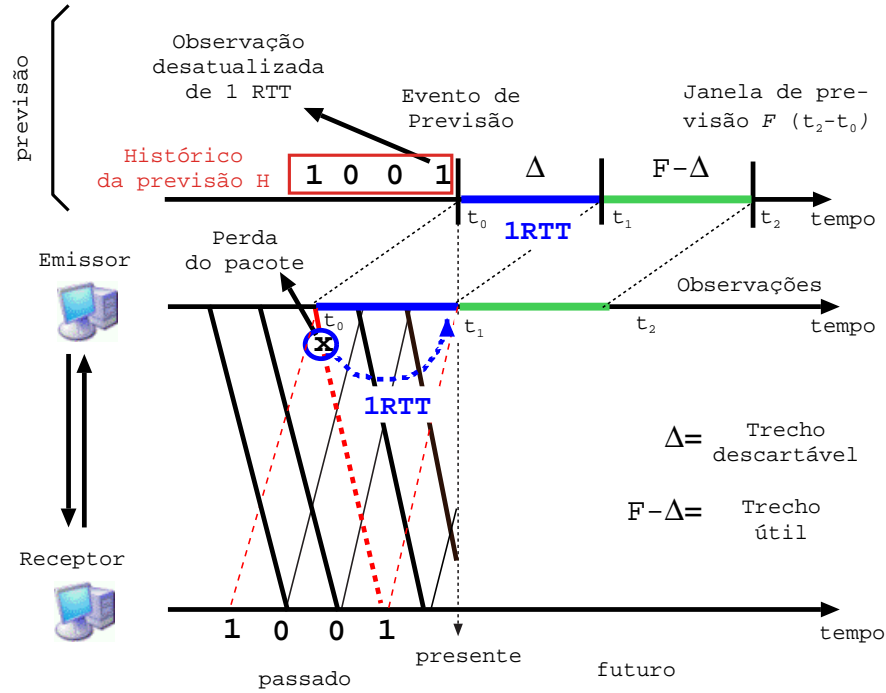


Figura 4.11: Atraso das observações em um dado evento de previsão

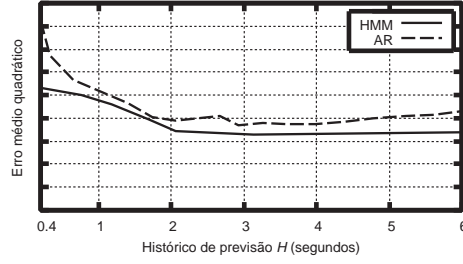


Figura 4.12: MSE do mecanismo de previsão adaptativa variando o histórico da previsão  $H$  para o *trace* da Figura 4.9(a)

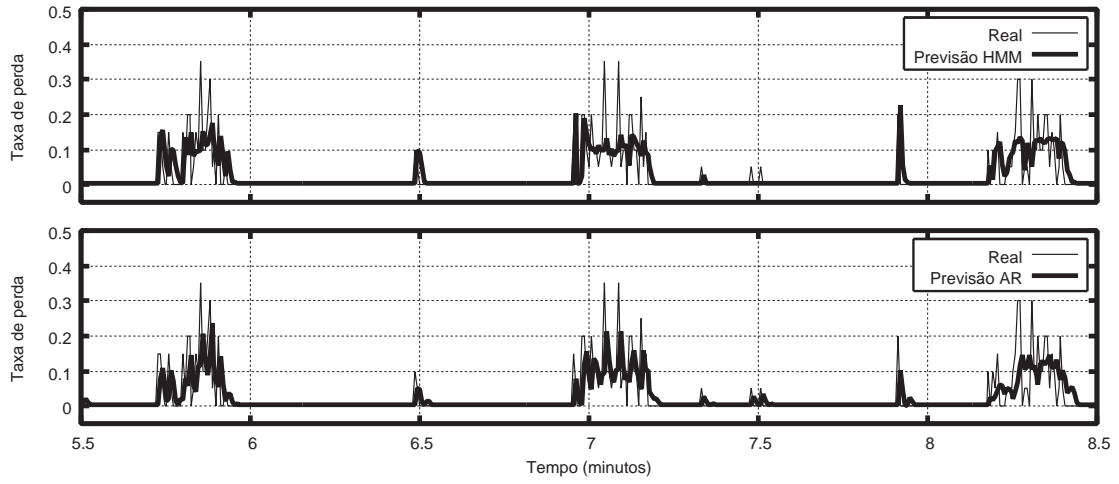


Figura 4.13: Previsão em uma medição entre UMd e UFRJ

Tabela 4.2: Comparação de medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos no *trace* da Figura 4.9(a),  $H = 2$  segundos e  $T = 2$  minutos

Modelo	MSE (Variabilidade)	COR (Variabilidade)	$I(0.4)$ (Variabilidade)
HMM	0.01244	0.23599	10%
AR	0.01330	0.22376	9%
Replicador	0.01823	0.16268	0%
Média	0.02189	0.10525	1%

A Figura 4.12 mostra que um  $H > 2$  segundos é irrelevante para o processo de previsão. Com  $H = 2$  segundos, os resultados do primeiro *trace* mostram que o modelo AR e o modelo HMM tiveram um desempenho semelhante. A Figura 4.13

sugere que o AR é melhor para representar a variabilidade da taxa de perda. No entanto, o modelo AR está apenas “replicando” as observações passadas.

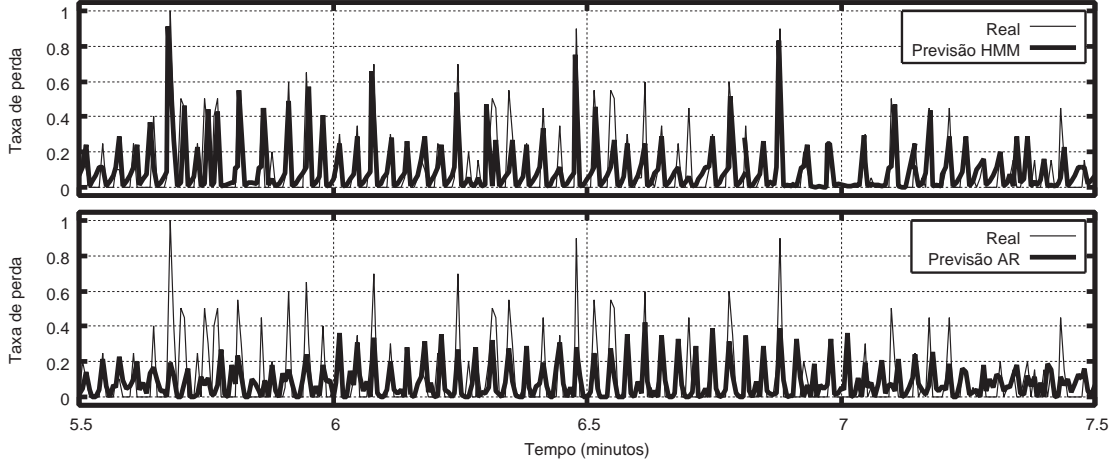


Figura 4.14: Previsão em uma medição entre UMass e UFRJ

Tabela 4.3: Comparação de medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos no *trace* da Figura 4.9(b),  $H = 0.8$  segundos e  $T = 10$  segundos

Modelo	MSE (Variabilidade)	COR (Variabilidade)	$I(0.4)$ (Variabilidade)
HMM	0.04599	0.43518	25%
AR	0.05625	0.41024	16%
Replicador	0.11260	-0.16332	0%
Média	0.08989	0.00520	0%

No segundo *trace*, o modelo HMM e o modelo AR tem um desempenho geral semelhante. Como neste caso existe uma estrutura autorregressiva vista na periodicidade da taxa de perda, o modelo AR terá um desempenho geral comparável ao modelo HMM.

Se olharmos a Figura 4.14, vemos que o HMM tem um desempenho superior para representar os picos de perda. Por isto, nas colunas de variabilidade da Tabela 4.3, o HMM é significativamente melhor.

Neste caso, também escolhemos o valor do histórico de previsão  $H$  avaliando sua influência no MSE.

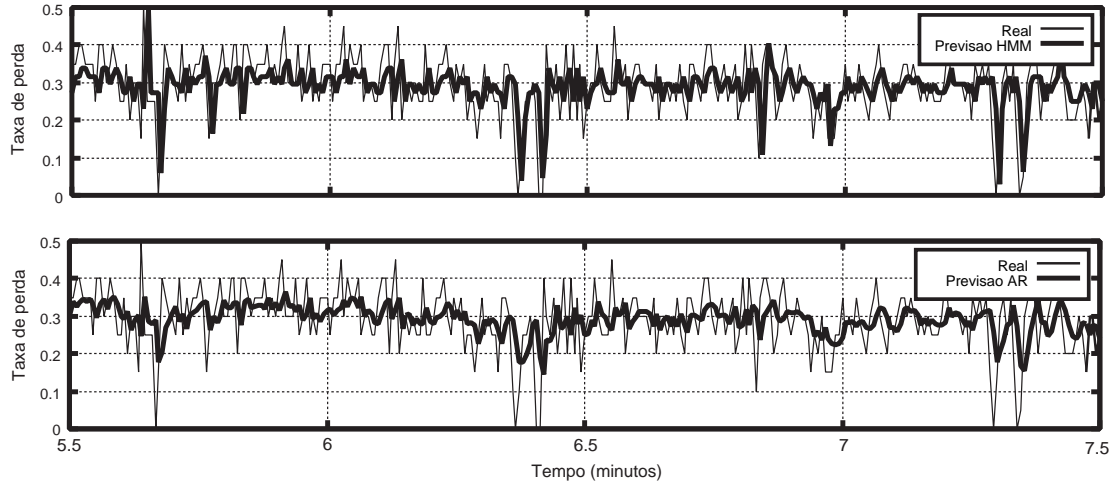


Figura 4.15: Previsão em uma medição entre UFMG e UFRJ

Tabela 4.4: Comparação de medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos no *trace* da Figura 4.9(c),  $H = 800$  milissegundos e  $T = 1$  minuto

Modelo	MSE (Variabilidade)	COR (Variabilidade)	$I(0.4)$ (Variabilidade)
HMM	0.00782	0.26941	16%
AR	0.01039	0.16756	14%
Replicador	0.01373	0.12242	0%
Média	0.01511	0.15834	10%

Pela Tabela 4.4, nota-se que a previsão com o modelo HMM e o modelo AR têm um MSE semelhante. Entretanto, novamente nos períodos de variação, o HMM é melhor, como vemos na Figura 4.15.

Por fim, notamos que em todos os *traces*, o papel de  $\tau$  foi praticamente irrelevante porque estes *traces* não apresentavam mudanças significativas ao longo do tempo. A Figura 4.16 exemplifica o MSE para o HMM da Tabela 4.4 variando o intervalo de reestimação  $\tau$ . Na Figura também vemos o MSE de um modelo HMM sem reestimações ao longo de todo o *trace*.

Em resumo, os experimentos com *trace* mostraram que nos períodos de variação, o HMM é melhor. E em todos os *traces*, o algoritmo de previsão foi insensível ao

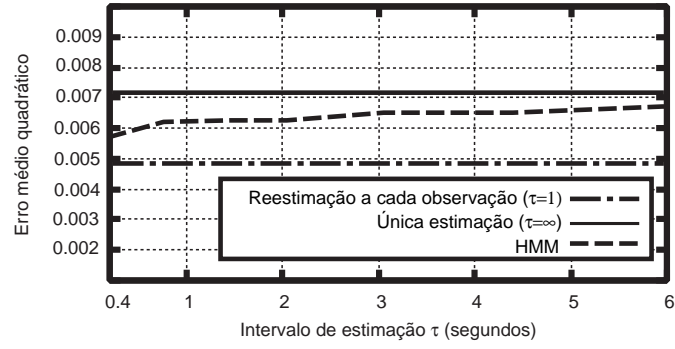


Figura 4.16: MSE do mecanismo de previsão adaptativa variando o intervalo de estimação da previsão  $\tau$  para o *trace* da Figura 4.9(c) com  $H = 800$  milisegundos e  $T = 1$  minuto (3000 amostras)

intervalo entre reestimões do modelo  $\tau$ .

## 4.5 Cenários simulados

Nesta seção, avaliaremos a capacidade de previsão e a sensibilidade dos parâmetros do mecanismo de previsão adaptativa em cenários simulados. O objetivo é avaliar o mecanismo em situações controladas.

Nestes cenários, a aplicação VoIP simplificada descrita na Seção 4.4 e o tráfego de fundo compartilharão um único recurso: um roteador de capacidade  $C$  *Mbps* e *buffer* de  $B$  pacotes (Figura 4.17). Escolhemos esta topologia com apenas um garagalo porque desejamos prever a variabilidade da taxa de perdas no *buffer*.

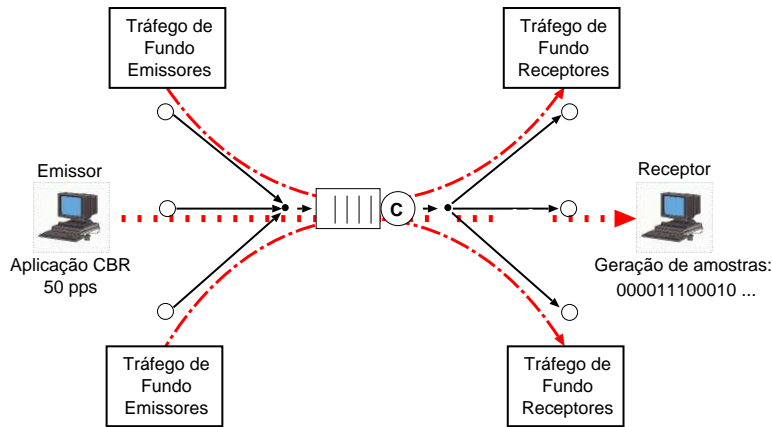


Figura 4.17: Topologia geral da simulação

Arbitramos o retardo de propagação de ambos os sentidos em 100 milissegundos, a capacidade do canal em  $C = 10\text{Mbps}$  ( $\gg 140.80\text{ Kbps}$ ) e o *buffer* em  $B = 100$  pacotes. Para obter os resultados utilizaremos o ambiente de simulação de redes de computadores **NS-2** [McCanne e Floyd 2007].

Selecionamos dois tipos de tráfego de fundo; no primeiro, temos um conjunto de fluxos TCP de longa e curta duração. Escolhemos este tipo de tráfego devido a predominância de tráfego TCP na *Internet*. No segundo, utilizamos um tráfego “auto-similar”. Este tráfego representaria a agregação de diversos tipos de fluxos multimídia (e.g. *web*, vídeo, voz).

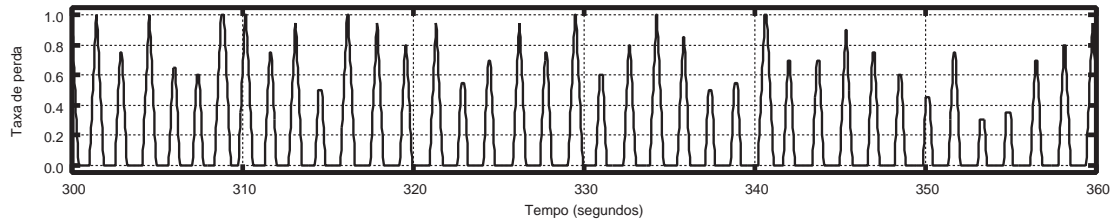


## Tráfego de fundo TCP

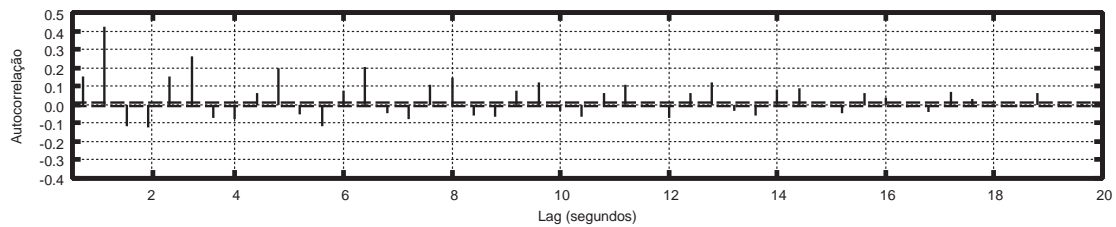
Os resultados deste experimento foram obtidos simulando a aplicação VoIP simplificada concorrendo com 40 fontes TCP's de duração "infinita" de mesmo RTT. Como a taxa de transmissão da aplicação é constante, as perdas observadas no roteador estarão agrupadas em rajadas devido à sincronização da dinâmica das taxas entre os fluxos TCP's [Shenker et al. 1990].

Para gerar ruído aos fluxos, a intervalos de tempo aleatórios com distribuição exponencial de média 60 segundos, 100 fontes TCP's transmitiam poucos 60 KB, exigindo uma banda agregada média de 800 Kbps.

A Figura 4.18(a) mostra como a taxa de perda observada pela aplicação VoIP é periódica em consequência da sincronização dos algoritmos de controle de congestionamento do TCP. Através da Figura 4.18(b), observamos uma estrutura recorrente na função de autocorrelação da taxa de perda a cada 2 segundos.



(a) Taxa de perda ao longo do tempo



(b) Autocorrelação da taxa de perdas

Figura 4.18: Taxa de perda e autocorrelação

Antes de obtermos os resultados do mecanismo de previsão, é necessário estimar

o número de estados do modelo de Markov oculto. A metodologia da Seção 2.2.2 estimou o número de estado em  $N = 1$ . Entretanto, este valor não é coerente com o processo de perdas porque o intervalo entre perdas não tem distribuição geométrica. Isso indica também que a entropia pode não ser um bom estimador para o número de estados em todos os casos.

Decidimos utilizar  $N = 20$  estados com a estrutura da Figura 4.19, onde cada estado representa 100 milissegundos do processo de perdas, totalizando os 2 segundos da estrutura periódica observada na taxa de perda.

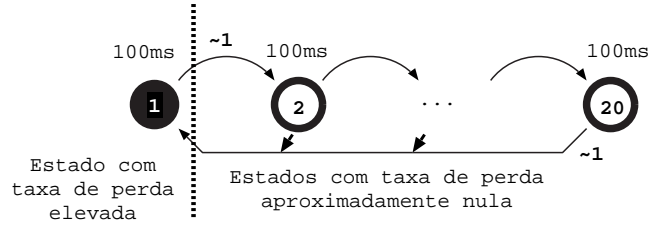


Figura 4.19: Cadeia com estrutura particular *coxian*

Para este processo de perda, o intervalo entre eventos de estimação  $\tau$  não é relevante porque o processo de perdas é estacionário. Experimentalmente, temos que com  $T > 1$  minuto (3000 amostras) a distribuição da fração média de perda e a rajada média de perdas amostrais convergem para as respectivas distribuições amostrais. Em relação à previsão, na Figura 4.20, vemos o MSE do mecanismo de previsão adaptativa em função do histórico de previsão  $H$ . Já a Tabela 4.5 resume as medidas de desempenho.

Tabela 4.5: Comparação de medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos,  $H = 2$  segundos e  $T = 1$  minuto

Modelo	MSE Variante	COR Variante	$I(0.4)$ Variante
HMM	0.07013	0.60975	14%
AR	0.08238	0.49841	11%
Replicador	0.24931	-0.23934	0%
Média	0.11135	0.04287	6%

A Figura 4.20 mostra que  $H > 2$  ( $p > 5$ ) segundos não diminuirá considera-

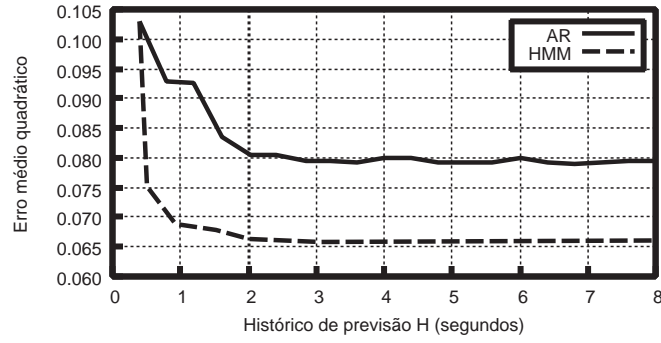


Figura 4.20: MSE do mecanismo de previsão adaptativa variando o histórico da previsão  $H$  com uma janela de previsão de  $400ms$

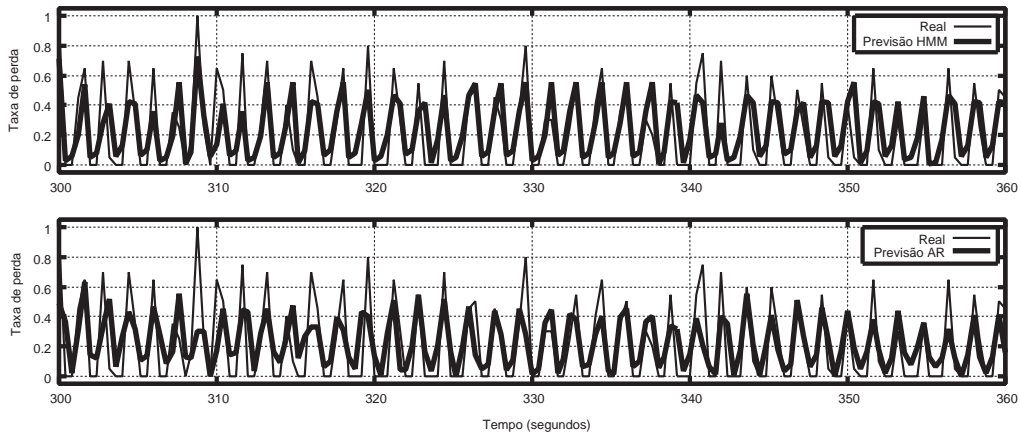


Figura 4.21: Previsão em um ambiente simulado com tráfego TCP

velmente o MSE da previsão com o modelo HMM ( $AR(p)$ ). Esta observação é consequência imediata do período médio da taxa de perda ser de 2 segundos (Fig. 4.18).

Da Tabela 4.21, temos que o desempenho do modelo HMM, em termos do MSE e da correlação da previsão é ligeiramente melhor do que o modelo AR. Entendemos que esta diferença se deve à estrutura da cadeia oculta do HMM.

## Tráfego de fundo MMPP

Como foi dito anteriormente, a motivação deste experimento é simular um ambiente com tráfego de aplicações multimídias. A principal característica deste tipo de tráfego, é observada na auto-similaridade da sua intensidade [Sahinoglu e Tekinay 1999].

Os resultados foram obtidos usando um processo MMPP (*Markovian Modulated Poisson Process*) como tráfego de fundo. O processo MMPP é simplesmente um processo Poisson com intensidade  $\lambda_t$ , sendo  $\{\lambda_t\}$  um processo markoviano. Pela proposta de [Robert e Le Boudec 1997] temos uma cadeia de Markov específica que nos permite gerar tráfego pseudo auto-similar a partir de apenas 3 parâmetros: a intensidade média do tráfego, o expoente de Hurst  $H$  e a “duração” da auto-similaridade.

Arbitramos a intensidade média em 20% da capacidade do canal e os demais parâmetros foram escolhidos iguais ao exemplo de maior variabilidade apresentado em [Robert e Le Boudec 1997]. A Figura 4.23 mostra a variabilidade do tráfego MMPP arbitrado e do tráfego Poisson de mesma intensidade média em diferentes escalas de tempo.

A variabilidade na intensidade do tráfego apresentada em escalas de tempo elevadas ocasiona perdas por congestionamento com uma correlação temporal maior, como vemos na Figura 4.22, do que em outros modelos de tráfego como o Poisson por exemplo. A Figura 4.26(a) mostra a variabilidade da taxa de perda para este cenário.

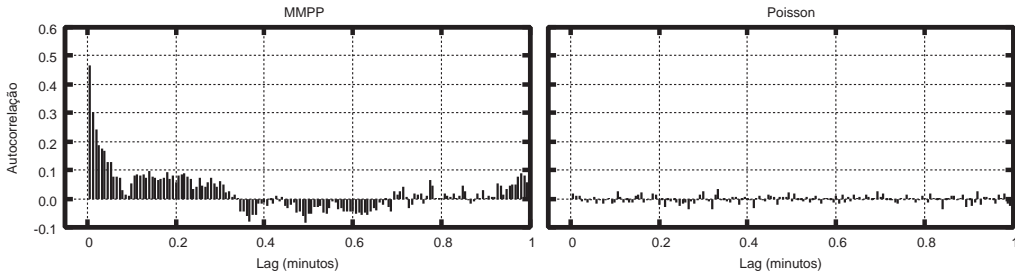


Figura 4.22: Autocorrelação da taxa de perdas

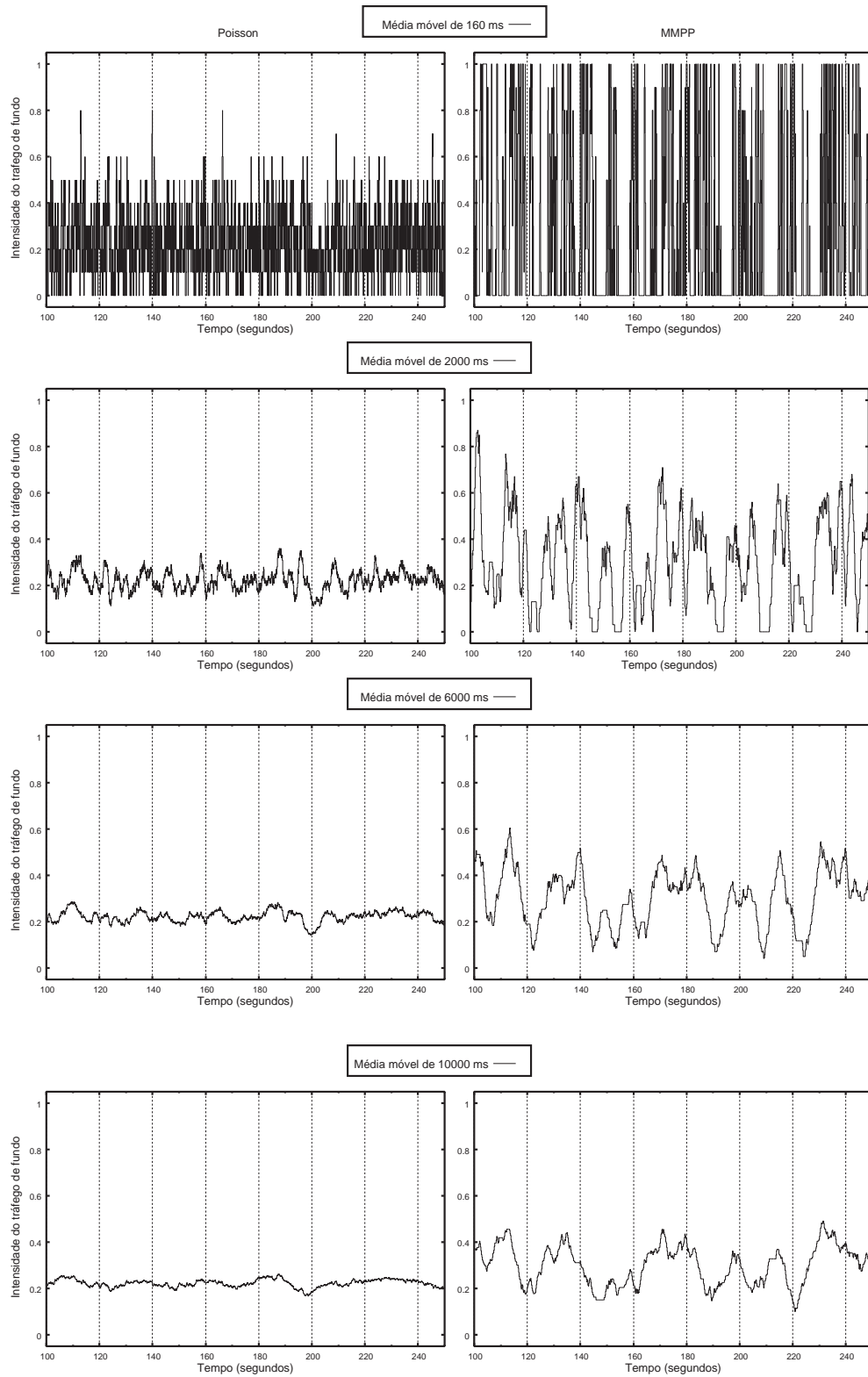
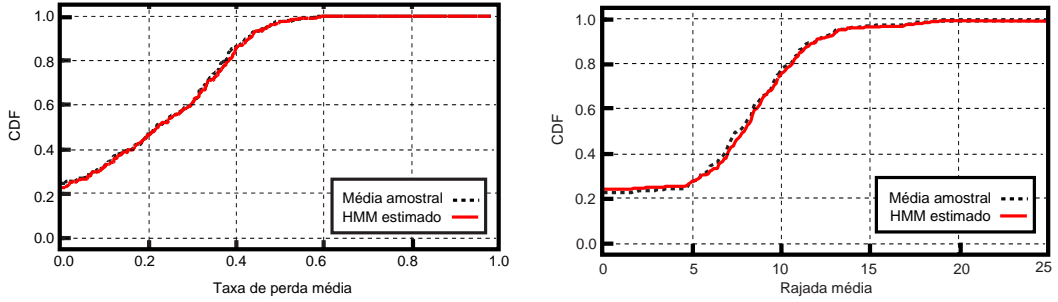


Figura 4.23: Intensidade de tráfego em quatro diferentes escalas de tempo para um processo Poisson e um processo MMPP ambos com mesma média

Pela metodologia usual, o número de estados estimado para o modelo HMM foi  $N = 5$ . Neste caso, um histórico de treinamento de  $T = 500$  foi suficiente para que a distribuição da fração média de perda e da rajada média de perdas estimadas no modelo HMM,  $\hat{\mu}_X$  e  $\hat{\mu}_Z^T$ , convergissem para as respectivas distribuições amostrais - Figura 4.24. Enquanto o erro dos estimadores amostrais torna-se constante a partir de  $T = 1000$  amostras do processo de perdas.



(a) Comparação entre a distribuição de  $l_T$  e a de  $\hat{\mu}_X$  obtidas empiricamente

(b) Comparação entre a distribuição de  $r_T$  e a de  $\hat{\mu}_Z$  obtidas empiricamente

Figura 4.24: Convergência da HMM de 5 estados

Na Figura 4.25 temos o MSE da previsão utilizando o modelo HMM e o modelo AR. Na Tabela 4.6 temos as medidas de desempenho do preditor com os modelos considerados. Uma amostra da previsão encontra-se na Figura 4.26(b).

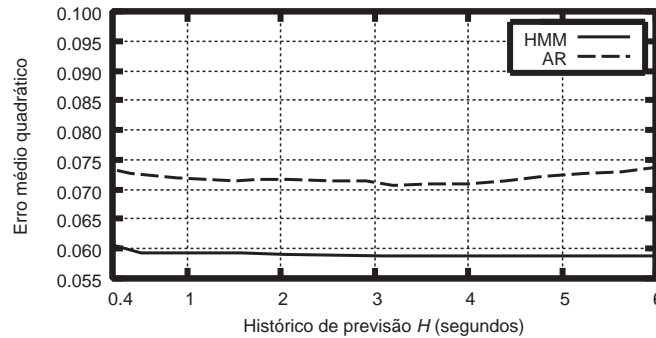


Figura 4.25: MSE do mecanismo de previsão adaptativa variando o histórico da previsão  $H$  com uma janela de previsão de  $400ms$

A Figura 4.25 mostra que um histórico de previsão  $H > 0.4 = F$  ( $p \geq 1$ )

Tabela 4.6: Comparação de medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos,  $H = 400$  milissegundos ( $p=1$ ) e  $T = 1$  minuto

Modelo	MSE Variante	COR Variante	$I(0.4)$ Variante
HMM	0.09548	0.24648	5%
AR(1)	0.11020	0.07927	6%
Replicado	0.17261	0.09643	0%
Média	0.21204	0.03003	6%

segundos não diminui o MSE do HMM (AR). De fato, a taxa de perda está muito correlacionada com seu último valor ( $\rho_R(1) = 0.84$ ). Conseqüentemente, a sua autocorrelação é significativa para  $h > 0$ . Através da função de autocorrelação parcial\*, vemos que a correlação relevante da taxa de perda é da ordem de 0.4 segundos.

Já os valores da Tabela 4.6 mostram que ambos os modelos HMM e AR têm uma taxa de acerto próxima de um preditor aleatório ( $1/(F + 1) \sim 5\%$ ) ou menor porque não conseguem prever a taxa de perdas corretamente. Pela Figura 4.26(b), vemos que a previsão está deslocada no tempo.

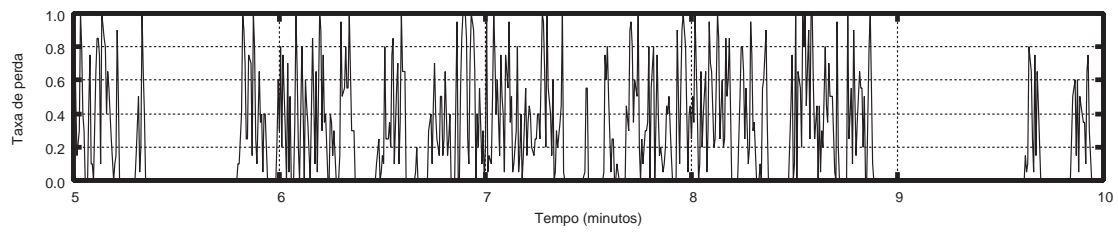
Vejamos o caso do modelo AR(1) apresentado na tabela. A variação entre a última observação e a previsão consecutiva é

$$\begin{aligned}
 |\hat{R}_t^F - R_{t-1}^F| &= |\rho(1)(R_{t-1}^F - \mu_R) + \mu_R - R_{t-1}^F| \\
 &= |(1 - \rho(1))| \cdot |(R_{t-1}^F - \mu_R)| = 0.16|(R_{t-1}^F - \mu_R)|.
 \end{aligned}$$

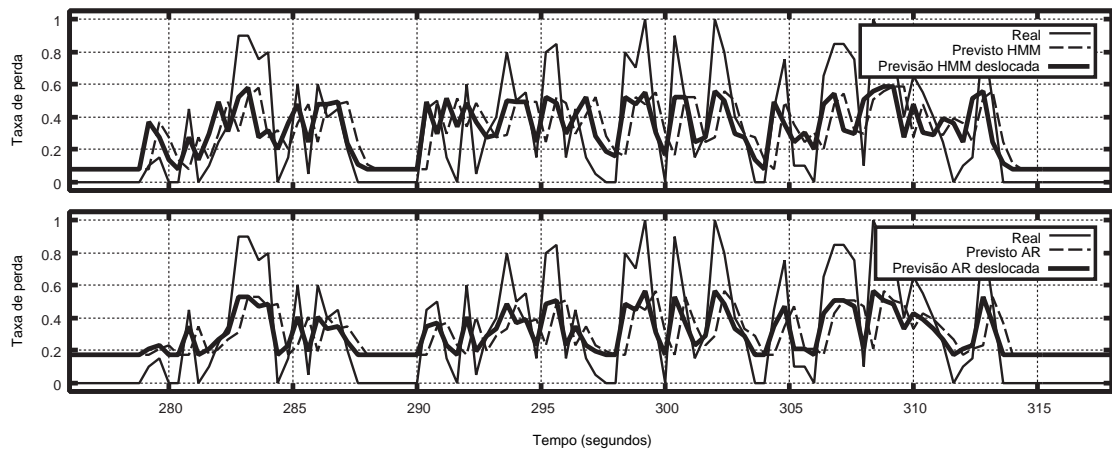
Conseqüentemente, em instantes de tempo consecutivos, teremos uma previsão “replicadora” escalonada em  $|(1 - \rho(1))| = 0.16$ .

---

\*A função autocorrelação parcial  $\alpha(i)$  mede a correlação de  $X_t$  com  $X_{t-i}$  não causada pelas correlações intermediárias[Brockwell e Davis 2002].



(a) Taxa de perda com alta variabilidade em um cenário simulado com tráfego de fundo MMPP



(b) Previsão em um ambiente simulado com tráfego MMPP

Figura 4.26: Taxa de perda e previsão em um cenário simulado com tráfego de fundo MMPP



## 4.6 Resumo dos experimentos

Em resumo, destacamos as seguintes observações dos experimentos.

1. Na maioria dos casos considerados, os modelos HMM e AR foram muito melhores do que os preditores replicador, média e aleatório em termos de erro médio quadrático, correlação cruzada e taxa de acerto. Além disto, o modelo HMM foi tão ou mais bem sucedido em acertar a taxa de perda durante os segmentos com certa variabilidade do que o modelo AR.
2. A respeito do histórico de previsão  $H$ , observamos que quando não existe um evento periódico, seu valor é relativamente pequeno.

Experimentalmente, vimos que o número de amostras  $T$  para estimação do modelo tem pouca influência no desempenho do preditor em processos estacionários. No entanto, não relacionamos formalmente o erro de estimação com o erro de previsão.

A janela de previsão é o parâmetro mais importante porque tem uma influência nítida na capacidade de previsão do algoritmo. Por outro lado, ela será determinada de acordo com a aplicação do preditor.

3. De acordo com a Tabela 4.7,  $T = 1$  minuto é suficiente para a estimação do modelo. A exceção ocorre quando existe algum evento periódico de longo prazo – como no *trace* UMd/UFRJ.

Considerando uma janela de previsão  $F = 400$  milissegundos,  $H = 1$  segundo seria um valor razoável caso não haja um evento periódico com intervalos maiores do que este 1 segundo.

A relação dos parâmetros e desempenho do algoritmo adaptativo em cada experimento segue na Tabela 4.7.

Tabela 4.7: Relação de parâmetros e medidas de desempenho para uma janela de previsão  $F = 400$  milissegundos

	HMM	AR	Replicador	Média
Modelo	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
Gilbert-Elliot	28%	13%	5%	1%
	<b>MSE</b> = 0.03281	<b>MSE</b> = 0.04367	<b>MSE</b> = 0.05761	<b>MSE</b> = 0.06632
	$H = 1$ pacote	$p = 1$		
	$T = 3000$ pacotes	$T = 3000$ pacotes		
<i>Trace</i>	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
(UMd/UFRJ)	10%	9%	0%	1%
	<b>MSE</b> = 0.00270	<b>MSE</b> = 0.00277	<b>MSE</b> = 0.00376	<b>MSE</b> = 0.00572
	$H = 2s$	$H = 2.4s(p = 6)$		
	$T = 2min$	$T = 2min$		
<i>Trace</i>	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
(UMass/UFRJ)	25%	16%	0%	0%
	<b>MSE</b> = 0.02539	<b>MSE</b> = 0.02712	<b>MSE</b> = 0.06339	<b>MSE</b> = 0.04144
	$H = 0.8s$	$H = 0.8s(p = 2)$		
	$T = 10s$	$T = 10s$		
<i>Trace</i>	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
(UFMG/UFRJ)	16%	14%	0%	10%
	<b>MSE</b> = 0.00610	<b>MSE</b> = 0.00667	<b>MSE</b> = 0.01017	<b>MSE</b> = 0.00699
	$H = 0.8s$	$H = 0.8s(p = 2)$		
	$T = 60s$	$T = 60s$		
Simulação	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
TCP	14%	11%	0%	6%
	<b>MSE</b> = 0.06369	<b>MSE</b> = 0.07894	<b>MSE</b> = 0.21077	<b>MSE</b> = 0.10310
	$H = 2s$	$H = 2s(p = 6)$		
	$T = 60s$	$T = 60s$		
Simulação	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$	Variante $I(0.4) =$
MMPP	5%	6%	0%	6%
	<b>MSE</b> = 0.05758	<b>MSE</b> = 0.07043	<b>MSE</b> = 0.09543	<b>MSE</b> = 0.09764
	$H = 0.4s$	$H = 0.4s(p = 1)$		
	$T = 60s$	$T = 60s$		

## Capítulo 5

# Considerações finais e trabalhos futuros

Nesta dissertação, continuamos com o estudo do algoritmo adaptativo proposto em [Silveira e de Souza e Silva 2006]. Neste sentido, avaliamos a capacidade de previsão do algoritmo adaptativo em cenários mais gerais do que os previamente considerados e caracterizamos a influência dos parâmetros do algoritmo adaptativo no seu desempenho.

Como contribuição fundamental deste trabalho temos o desenvolvimento de uma metodologia para avaliação das previsões baseada em técnicas estatísticas. Como resultado desta avaliação, foi possível mostrar que o algoritmo é robusto a variações de seus parâmetros.

Além disto, propomos a utilização de um modelo preditivo autorregressivo comparativamente ao modelo HMM. Embora o modelo AR tenha um desempenho razoável, veremos que o modelo HMM é superior em vários cenários e representa melhor a variabilidade da taxa de perdas.

Este trabalho também contribui apresentando uma revisão de diversas técnicas matemáticas aplicadas nos problemas relacionados direta ou indiretamente com

análise preditiva.

## Trabalhos futuros

As principais questões que surgiram ao longo desta dissertação que poderiam ser analisadas em um trabalho subsequente são

1. Analisar a relação entre o erro de previsão com erro de estimação do modelo HMM. O ponto de partida para esta análise seria quantificar o erro da convergência de algumas estatísticas transientes do modelo HMM para as do processo de perdas original. Ao longo deste trabalho, a estimação do modelo foi validada pela taxa de perda média e rajada de perda de estado estacionário.
2. Avaliar outras estatísticas para a previsão no modelo HMM, e.g.  $\hat{R}_t^F = \arg \max P[R_t^F | X_H, \dots, X_{t-1}]$ . A motivação inicial de utilizar o preditor ótimo era minimizar o MSE. No entanto, isto não implica que aumentaremos a taxa de acerto do preditor.

Há também a possibilidade de realizar estimação da distribuição probabilística do vetor  $R_t^{F'}, R_{t+F'}^{F'}, \dots, R_{t+kF'}^{F'}$ , onde  $kF' = F$ , condicionada às observações  $X_{t-H}, \dots, X_{t-1}$ . Isto nos daria uma granularidade menor para a previsão. Notemos que esta abordagem é essencialmente diferente de realizar uma nova previsão a cada  $F/k$  segundos.

3. Utilizar o algoritmo de previsão em um controle de congestionamento eficiente baseado em alguma heurística. Por exemplo, obter uma taxa de transmissão que seja a mais adequada às condições de perdas previstas para uma janela no futuro.
4. Avaliar o impacto do erro de previsão de diferentes modelos preditivos (e.g.: AR, HMM) em uma determinada aplicação.

# Apêndice A

## Conceitos teóricos adicionais

Este apêndice resume alguns conceitos teóricos que são utilizados ao longo da dissertação. O motivo para estarem separados é não tornar os capítulos anteriores muito extensos.

### A.1 Erro médio quadrático e o preditor ótimo

Seja  $X$  uma variável aleatória e  $\hat{X}$  uma previsão para uma realização qualquer de  $X$ . Uma forma de avaliar esta previsão, é através do erro médio quadrático, ou simplesmente **MSE** – *Mean Square Error*:

$$\mathbf{MSE} = E \left[ (X - \hat{X})^2 \right]. \quad (\text{A.1})$$

Condicionando o MSE a uma variável aleatória  $Y$ , temos que

$$\begin{aligned} \mathbf{MSE} &= E \left[ (X - \hat{X})^2 \right] \\ &= \sum_y E \left[ (X - \hat{X})^2 | Y = y \right] P[Y = y] \end{aligned}$$

Teremos o mínimo desta expressão, anulando a derivada em  $\hat{X}$ :

$$\sum_y E \left[ (X - \hat{X})^2 | Y = y \right] P[Y = y] = 0$$

Como  $P[Y = y]$  é não negativa para todo  $y$  e positiva para algum  $y$ , temos o erro médio quadrático mínimo em  $\hat{X} = E[X|Y]$ . Denominamos  $\hat{X} = E[X|Y]$  por **preditor ótimo** em consideração à  $Y$ ; e  $Y$  por informação disponível para a previsão. O valor do MSE neste ponto de mínimo é

$$\begin{aligned} \text{MSE}^* &= E[(X - E[X|Y])^2] \\ &= \sum_y E[(X - E[X|Y = y])^2 | Y = y] P[Y = y] \\ &= E[X^2] - \sum_y E[X|Y = y]^2 P[Y = y], \end{aligned}$$

Se  $X$  e  $Y$  são independentes,  $\hat{X} = E[X|Y] = E[X]$  e o MSE se reduz à variância  $V[X]$  e  $Y$  é inútil na previsão. Caso contrário,  $Y$  passa a atuar na previsão de  $X$  e o MSE é menor do que  $V[X]$ . Desta forma, o erro do **preditor ótimo** fica limitado entre

$$0 \leq \overset{\text{dependência}}{\underset{\text{de } X \text{ e } Y}{\leftarrow}} \text{MSE do preditor ótimo} \leq \overset{\text{independência}}{\underset{\text{de } X \text{ e } Y}{\rightarrow}} V[X]$$

Este resultado é generalizável para  $\hat{X} = E[X|Y_1, \dots, Y_H]$ , sendo  $Y_1, \dots, Y_H$  o conjunto de informação disponível para a previsão. Neste caso, o preditor ótimo tem um MSE de

$$\begin{aligned} \text{MSE}^* &= E[(X - E[X|Y_1, \dots, Y_H])^2] \\ &= E[X^2] - \sum_{y_1, \dots, y_H} E[X|Y_1, \dots, Y_H]^2 P[Y_1, \dots, Y_H]. \end{aligned} \quad (\text{A.2})$$

## A.2 Modelos de Gilbert-Elliot

No modelo de Gilbert-Elliot [Gilbert 1960, Elliott 1965], o estado do canal é representado por uma cadeia de Markov de tempo discreto, ou contínuo, de 2 estados - Figura A.1. Em cada estado, um pacote é descartado ou corrompido com uma certa probabilidade constante ao longo do tempo. Sem perda de generalidade, se  $l_b < l_r^*$  podemos qualificar os estados como bom e ruim respectivamente.

---

\*Se  $l_b = l_r$ , temos um processo de Bernoulli.

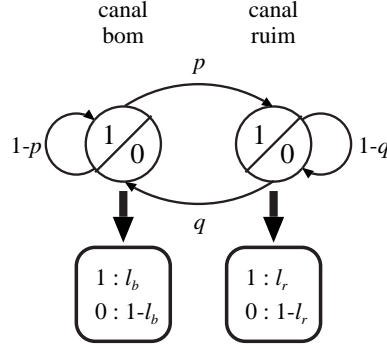


Figura A.1: Modelo de Gilbert-Elliot de tempo discreto

Seja  $\{X_t\}$  o processo estocástico de observações que representam os eventos de perda –  $X(t) = 1$  ( $0$ ) representa um pacote descartado (recebido) – e  $\{Y(t)\}$  o estado do canal de acordo com o modelo de Gilbert-Elliot. Este par de processos é um HMM como segue

1. Número de estados  $N = 2$  (bom e ruim).
2. Número de observações  $M = 2$  por estado ( $X(t) \in \{0, 1\}$ ).
3. Matriz de transição,

$$A = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

4. Matriz de distribuição de probabilidades das observações de um estado,

$$B = \begin{bmatrix} 1-l_b & l_b \\ 1-l_r & l_r \end{bmatrix}$$

5. Vetor de distribuição do estado inicial  $\pi = [\pi_b \ \pi_r]$ .

No caso onde no estado bom (ruim) o pacote é descartado pela rede com probabilidade 0 (1), i.e.,  $l_b = 0$  e  $l_r = 1$ , denominamos esta instância particular de **Gilbert-Elliot simplificado**, como vemos na Figura A.2.

A distribuição estacionária do estado oculto é

$$\pi = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix}, \quad (\text{A.3})$$

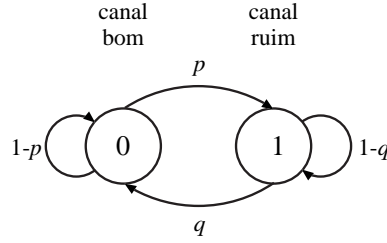


Figura A.2: Modelo de **Gilbert-Elliot simplificado** de tempo discreto

resultando em uma taxa média de erro  $E[X_t] = \pi \cdot [0 \ 1]' = \frac{p}{p+q}$ .

Seja  $I$  a variável aleatória que representa o número de transmissões bem sucedidas entre duas perdas e  $B$  do número de perdas consecutivas.  $I$  e  $B$  têm as respectivas distribuições geométricas

$$\begin{aligned} P[I = n] &= P[X_{t+n+1} = 1, X_{t+n} = 0, \dots, X_{t+1} = 0 | X_t = 1] \\ &= (1-p)^n p, \quad n \geq 0. \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} P[B = n] &= P[X_{t+n+1} = 0, X_{t+n} = 1, \dots, X_{t+1} = 1 | X_t = 0] \\ &= (1-q)^n q, \quad n \geq 0. \end{aligned} \quad (\text{A.5})$$

### A.3 Distribuição da média amostral

Freqüentemente desejamos obter a média de um conjunto de variáveis aleatórias  $X_1, \dots, X_n$ . Seja a média amostral definida pela estatística

$$\bar{\mu}_x = \frac{1}{n} \sum_{i=1}^n X_i. \quad (\text{A.6})$$

Se este conjunto é independente e identicamente distribuído com média  $E[X_i] = \mu_X$  e variância  $Var[X_i] = \sigma_X^2$  conhecidas, pelo Teorema Central do Limite temos que  $\bar{\mu}_X \sim N(\mu_X, \sigma_X^2/n)$ . Assim, podemos derivar um intervalo de confiança em torno da média real  $\mu_X$  para o estimador.

A situação se torna um pouco mais complicada se as variáveis apresentam alguma dependência. Consideremos apenas o caso particular onde  $X_1, \dots, X_n$  formam um



processo estocástico estacionário de segunda ordem com média  $\mu_X$  e dependência descrita pela função de autocorrelação  $\rho_X(h)$  ( $Var[X_i] = \rho(0)$ ). É imediato observar que o estimador amostral continua sendo consistente e não tendencioso, entretanto a sua variância não é  $\sigma_X^2/n$ . De fato, temos que

$$Var[\bar{\mu}_X] = \frac{1}{n} \left( \sigma^2 + 2 \sum_{j=1}^{n-1} (1 - j/n) \rho(j) \right), \quad (A.7)$$

se  $n \gg 0$ , a expressão acima pode ser aproximada por

$$Var[\bar{\mu}_X] = \frac{1}{n} \sigma^2 K, \quad K = 1 + 2 \sum_{j=1}^n \frac{\rho(j)}{\sigma^2}. \quad (A.8)$$

Sob condições gerais, temos que  $\bar{\mu}_X \sim N(\mu_X, K\sigma_X^2/n)$  para  $n$  suficientemente grande. Ou seja, a medida que as variáveis apresentam dependência, a variância do estimador aumenta.

De uma maneira geral, o intervalo de confiança  $100\%(1 - \alpha)$  da média amostral é  $\mu_X \pm \sigma z_{\alpha/2} \sqrt{\frac{K}{n}}$  (se as variáveis são independentes  $K = 1$ ).

A fração  $\frac{n}{K}$  é frequentemente denominada de tamanho efetivo das amostras independentes no conjunto de  $n$  amostras dependentes pelas razões acima.

## A.4 Distribuição da função de autocorrelação amostral - SACF

Seja  $\{Y(t)\}$  um processo independente e identicamente distribuído com média nula e variância  $\sigma^2 < \infty$ , denotado por  $\text{IID}(0, \sigma^2)$ . Para um número  $n$  grande de amostras a função de autocorrelação amostral  $\hat{\rho}_Y(h)$ ,  $|h| < n$ , se aproxima a uma variável aleatória normal  $N(0, 1/n)$ . Desta forma, cerca de 95% das autocorrelações estarão entre o intervalo  $[-1.96/\sqrt{n}, +1.96/\sqrt{n}]$ .

O seguinte teste de hipótese pode ser usado para verificar se um conjunto de dados é independente:

1. Hipótese nula  $H_0$ :  $\{Y(t)\}$  é um processo independente e identicamente distribuído com média nula e variância  $\sigma^2$ ;
2. Estatística de tese: Para um conjunto de  $n$  amostras de  $\{Y(t)\}$ , obter a SACF  $\hat{\rho}_Y(h)$ ,  $|h| < n$ ;
3. Rejeitar a hipótese com significância 0.05 se  $P[|\hat{\rho}_Y(h)| \geq 1.96/\sqrt{n}] \geq 0.05$ .

Outro teste aplicado para verificar se uma SACF é nula, i.e., o processo é IID( $0, \sigma^2$ ), é o teste de portmanteau onde a estatística de teste é  $Q = n(n + 2) \sum_{j=1}^h \hat{\rho}_Y^2(j)/(n - j) \sim \chi^2(h)$  e se rejeita  $H_0$  com significância  $\alpha$  se  $Q > \chi_{1-\alpha}^2(h)$ .

## A.5 Distribuição multinormal

**Definição A.1.** *Seja  $\mathbf{X} = (X_1, \dots, X_n)'$  um vetor de variáveis aleatórias,  $\mathbf{X}$  é multinormal com média  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$  e matriz de covariância  $\Sigma = \Sigma_{\mathbf{X}\mathbf{X}} = [\text{Cov}(X_i, X_j)]_{i,j}$ , denotado por  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , se a sua distribuição probabilística conjunta segue*

$$f_{\mathbf{X}}(\mathbf{x}) = P[\mathbf{X} = \mathbf{x}] = (2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}} \exp \left\{ \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (\text{A.9})$$

Ressaltamos que uma variável multinormal não é um conjunto de  $n$  variáveis normais independentes, neste caso teremos  $\Sigma = \mathbb{I}$  e  $f_{\mathbf{X}}(\mathbf{x})$  será o produto da distribuição das  $n$  variáveis normais. Em [Brockwell e Davis 2002], é mostrado que a transformação linear  $\mathbf{Z} = \Sigma^{1/2}(\mathbf{X} - \boldsymbol{\mu})$  leva o vetor  $\mathbf{X}$  a  $n$  variáveis aleatórias independentes com distribuição  $N(0, 1)$ .

Uma propriedade de interesse das variáveis multinormais é que a distribuição condicional de qualquer subconjunto  $\mathbf{X}^{(1)}$  de  $\mathbf{X}$  dado o complementar  $\mathbf{X}^{(2)}$  é também multinormal com distribuição  $N(\boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ , onde  $\Sigma_{ij} = [(\mathbf{X}^{(i)} - \boldsymbol{\mu}^{(i)})(\mathbf{X}^{(j)} - \boldsymbol{\mu}^{(j)})']$ . Com isto,  $E[\mathbf{X}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}] = \boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$ .

# Referências Bibliográficas

- [Bendat e Piersol 1971] Bendat, J. S. e Piersol, A. G. (1971). *Random data: analysis and measurement procedures*. Willey-Interscience, 3<sup>a</sup> edição.
- [Bolot et al. 1999] Bolot, J. C., Parisi, S. F., e Towsley, D. F. (1999). Adaptive FEC-based error control for Internet telephony. In *Proceedings of the IEEE INFOCOM*, pp. 1453–1460.
- [Boutremans e Boudec 2003] Boutremans, C. e Boudec, J. Y. L. (2003). Adaptive joint playout buffer and FEC adjustment for Internet telephony. *INFOCOM 2003*, 1:652–662.
- [Box et al. 1994] Box, G., Jenkins, G. M., e Reinsel, G. (1994). *Time Series Analysis: Forecasting & Control*. Prentice Hall, 2<sup>a</sup> edição.
- [Brakmo e Peterson 1995] Brakmo, L. S. e Peterson, L. L. (1995). TCP vegas: End to End congestion avoidance on a global Internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–1480.
- [Brockwell e Davis 2002] Brockwell, P. J. e Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer.
- [Cover e Thomas 2006] Cover, T. M. e Thomas, J. A. (2006). *Elements of Information theory*. John Wiley and Sons, 2<sup>a</sup> edição.
- [de Souza e Silva e Gail 2000] de Souza e Silva, E. e Gail, H. (2000). Transient solutions for Markov chains. *Computational Probability*, pp. 43–81.

- [de Souza e Silva et al. 2006] de Souza e Silva, E. A., da Silva, A. P. C., de A. Rocha, A. A., Leão, R. M. M., Duarte, F. P., Silveira, F., Jaime, G. D. G., e Muntz, R. R. (2006). Modeling, analysis, measurement and experimentation with the Tangram-II integrated environment. In *VALUETOOLS*, pp. 7.
- [Duarte 2003] Duarte, F. P. (2003). Algoritmo adaptativo para previsão e recuperação de perda de pacotes em aplicações multimídias usando cadeias de markov ocultas. Tese de Mestrado, Universidade Federal do Rio de Janeiro - COPPE - Programa de Engenharia de Sistemas e Computação.
- [Duarte et al. 2003] Duarte, F. P., de Souza e Silva, E., e Towsley, D. (2003). An adaptive FEC algorithm using hidden markov chains. *SIGMETRICS Performance Evaluation Review*, 31(2):11–13.
- [Elliot et al. 1995] Elliot, R. J., Aggoun, L., e Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control*. Springer-Verlag.
- [Elliott 1965] Elliott, E. O. (1965). A model of the switched telephone network for data communications. *Bell Systems Technical Journal*, 44:89–109.
- [Floyd et al. 2000] Floyd, S., Handley, M., Padhye, J., e Widmer, J. (2000). Equation-based congestion control for unicast applications. In *SIGCOMM*, pp. 43–56, Stockholm, Sweden.
- [FreeMeeting 2007] FreeMeeting (2007). Laboratory for modeling, ANalysis and Development of networks and computing systems. URL <http://www.land.ufrj.br>.
- [Gilbert 1960] Gilbert, E. N. (1960). Capacity of a burst-noise channel. *Bell Systems Technical Journal*, 39:1253–1265.
- [Ji et al. 2004] Ji, P., Liu, B., Towsley, D. F., Ge, Z., e Kurose, J. F. (2004). Modeling frame-level errors in GSM wireless channels. *Performance Evaluation*, 55(1-2):165–181.

- [Mascolo et al. 2001] Mascolo, S., Casetti, C., Gerla, M., Sanadidi, M. Y., e Wang, R. (2001). TCP westwood: Bandwidth estimation for enhanced transport over wireless links. In *Mobile Computing and Networking*, pp. 287–297.
- [McCanne e Floyd 2007] McCanne, S. e Floyd, S. (2007). NS-2 network simulator version 2.29. URL <http://www.isi.edu/nsnam/ns/>.
- [Merhav et al. 1989] Merhav, N., Gutman, M., e Ziv, J. (1989). On the estimation of the order of a Markov chain and universal data compression. *IEEE Transactions on Information Theory*, 35(5):1014–1019.
- [Paxson 1997] Paxson, V. (1997). End-to-end Internet packet dynamics. In *SIGCOMM '97: Conference proceedings on Applications, technologies, architectures, and protocols for computer communication*, pp. 139–152.
- [Perkins et al. 1998] Perkins, C., Hodson, O., e Hardman, V. (1998). A survey of packet loss recovery techniques for streaming audio. *Network IEEE*, 12(5):40–48.
- [Prezzi e Murai 2007] Prezzi, E. D. V. e Murai, F. (2007). Implementação e avaliação de um algoritmo de seleção Ótima de FEC baseado em cadeias de Markov ocultas para ferramentas de VoIP. Projeto Final de Curso, Universidade Federal do Rio de Janeiro - IM - Departamento de Ciência da Computação.
- [Rabiner 1989] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285.
- [Robert e Le Boudec 1997] Robert, S. e Le Boudec, J. Y. (1997). New models for pseudo self-similar traffic. *Performance Evaluation*, 30(1-2):57–68.
- [Ross 1982] Ross, S. M. (1982). *Stochastic Process*. John Wiley & Sons.
- [Sahinoglu e Tekinay 1999] Sahinoglu, Z. e Tekinay, S. (1999). On multimedia networks: self-similar traffic and network performance. *Communications Magazine, IEEE*, 37(1):48–52.

- [Salamatian e Vaton 2001] Salamatian, K. e Vaton, S. (2001). Hidden Markov modeling for network communication channels. In *Proceedings of the ACM SIGMETRICS*, pp. 92–101.
- [Shenker et al. 1990] Shenker, S., Zhang, L., e Clark, D. (1990). Some observations on the dynamics of a congestion control algorithm. *ACM Computer Communication Review*, pp. 30–39.
- [Silveira 2006] Silveira, F. (2006). Previsão de estatísticas de perdas de pacotes usando modelos de markov ocultos. Tese de Mestrado, Universidade Federal do Rio de Janeiro - COPPE - Programa de Engenharia de Sistemas e Computação.
- [Silveira e de Souza e Silva 2006] Silveira, F. e de Souza e Silva, E. A. (2006). Modeling the short-term dynamics of packet losses. *SIGMETRICS Performance Evaluation Review*, 34(3):27–29.
- [Silveira et al. 2006] Silveira, F., Watanabe, E. H., e de Souza e Silva, E. A. (2006). Adaptive Forward Error Correction for interactive streaming over the Internet. *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, pp. 1–6.
- [Su et al. 2004] Su, Y. C., Yang, C. S., e Lee, C. W. (2004). The analysis of packet loss prediction for Gilbert-model with loss rate uplink. *Information Processing Letters*, 90:155–159.
- [Tao e Guerin 2004] Tao, S. e Guerin, R. (2004). On-line estimation of Internet path performance: an application perspective. In *Proceedings of IEEE INFOCOM*, volume 3, pp. 1774–1785.
- [Trivedi 2001] Trivedi, K. (2001). *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. John Wiley and Sons, 2ª edição.
- [Yajnik et al. 1999] Yajnik, M., Moon, S. B., Kurose, J. F., e Towsley, D. F. (1999). Measurement and modeling of the temporal dependence in packet loss. In *Proceedings of the IEEE INFOCOM*, pp. 345–352.

[Ziv e Merhav 1992] Ziv, J. e Merhav, N. (1992). Estimating the number of states of a finite-state source. *IEEE Transactions on Information Theory*, 28:61–65.