

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Detecção de mudança de distribuição em dados sequenciais

Douglas Decicino de Andrade

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Douglas Decicino de Andrade

Detecção de mudança de distribuição em dados sequenciais

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Luis Ernesto Bueno Salasar

USP – São Carlos
Julho de 2025

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

D294d Decicino de Andrade, Douglas
 Detecção de mudança de distribuição em dados
 sequenciais / Douglas Decicino de Andrade;
 orientador Luis Ernesto Bueno Salasar. -- São
 Carlos, 2025.
 89 p.

 Dissertação (Mestrado - Programa
 Interinstitucional de Pós-graduação em Estatística) --
 Instituto de Ciências Matemáticas e de Computação,
 Universidade de São Paulo, 2025.

 1. Dataset shift. 2. Detecção de mudanças de
 distribuição. 3. Detecção de ponto de mudança. 4.
 Dados sequenciais. 5. Aprendizagem de máquina. I.
 Ernesto Bueno Salasar, Luis, orient. II. Título.

Douglas Decicino de Andrade

Distribution shift detection in sequential data

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.
FINAL VERSION

Concentration Area: Statistics

Advisor: Prof. Dr. Luis Ernesto Bueno Salasar

USP – São Carlos
July 2025



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Douglas Decicino de Andrade, realizada em 06/06/2025.

Comissão Julgadora:

Prof. Dr. Luis Ernesto Bueno Salasar (UFSCar)

Prof. Dr. Marcio Luis Lanfredi Viola (UFSCar)

Prof. Dr. George Lucas Moraes Pezzott (UEM)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

Este trabalho é dedicado aos meus pais e minha família, pilares que me sustentaram com seu apoio incondicional. Vocês foram a base que me inspirou a seguir em frente, mesmo diante dos desafios, e sua presença constante tornou possível a realização deste sonho. Sem vocês, nada disso seria possível.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer ao meu orientador, Prof. Dr. Luis Ernesto Bueno Salazar, por sua orientação e confiança durante todo o desenvolvimento desta pesquisa. Seu incentivo constante, disponibilidade e conhecimentos foram fundamentais para que este trabalho se tornasse realidade.

Agradeço, de modo muito especial, à minha família, que sempre esteve ao meu lado em cada etapa desta jornada. Ao meu pai, Geraldo, pelo exemplo de dedicação e força, e à minha mãe, Fátima, pelo carinho e compreensão incondicionais. Sem o apoio de vocês, muitas portas teriam permanecido fechadas, e muitos desafios teriam sido ainda maiores.

A vocês, minha gratidão eterna, por acreditarem nos meus sonhos, me incentivarem a buscar o melhor de mim e compartilharem cada conquista. Esse trabalho também é fruto do amor, do incentivo e da paciência que vocês sempre demonstraram.

Agradeço, por fim, a todos os familiares e amigos que, direta ou indiretamente, contribuíram para que eu pudesse chegar até aqui. Cada palavra de encorajamento, cada gesto de apoio e cada conselho recebido foram essenciais ao longo desta trajetória. Obrigado por fazerem parte desta conquista.

Por fim, agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES)- Código de Financiamento 001, pelo apoio financeiro sem o qual essa pesquisa não seria possível .

“In God we trust. All others must bring data.”
(W. Edwards Deming.)

RESUMO

ANDRADE, D. D. **Detecção de mudança de distribuição em dados sequenciais**. 2025. 89 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

Este trabalho se propõe a estudar o desempenho do método conformal para detecção de mudança de distribuição (*dataset shift*) para dados sequenciais, isto é, observações são sequencialmente obtidas ao longo do tempo. Em particular, consideramos os casos em que as observações são univariadas, multivariadas e informadas em lotes (*datasets*) em contextos preditivos. O método conformal gera valores- p a cada instante, que fornecem indícios a respeito de uma mudança na distribuição dos dados. Diferentemente de abordagens clássicas, como os métodos *Cumulative Sum Control Chart* (CUSUM), *Shirayev-Roberts* e o *Posterior Probability*, o método conformal não se baseia em suposições paramétricas sobre a distribuição dos dados antes e depois da mudança ocorrer. O método conformal foi aplicado considerando diferentes medidas de não conformidade que foram propostas neste trabalho, baseadas por exemplo na divergência de Kullback-Leibler e diferentes distâncias em espaços Euclidianos construídos a partir de medidas de profundidade estatística. A eficácia desta abordagem para detecção de mudança de distribuição foi avaliada em diferentes estudos empíricos por meio do tempo médio de atraso na detecção e pela proporção de falsos alarmes. Espera-se que esta pesquisa possa contribuir para a aplicação deste método no monitoramento *on-line* de modelos de aprendizado de máquina em ambientes dinâmicos, permitindo identificar o momento oportuno para recalibrar ou retreinar algoritmos de forma mais confiável.

Palavras-chave: *Dataset shift*, Detecção de mudanças de distribuição, Detecção de ponto de mudança, Dados sequenciais, Processamento em lote, Aprendizagem de máquina.

ABSTRACT

ANDRADE, D. D. **Distribution shift detection in sequential data**. 2025. 89 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

This work aims to study the performance of the conformal method for distribution change detection (*dataset shift*) in sequential data, where observations are obtained sequentially over time. In particular, we consider cases where the observations are univariate, multivariate, and provided in batches (datasets) in predictive contexts. The conformal method generates p -values at each time step, which provide evidence regarding a potential change in the data distribution. Unlike classical approaches, such as the *Cumulative Sum Control Chart* (CUSUM), *Shirayev-Roberts*, and *Posterior Probability* methods, the conformal method does not rely on parametric assumptions about the data distribution before and after the change occurs. The conformal method was applied using different nonconformity measures proposed in this work, including those based on Kullback-Leibler divergence and various distances in Euclidean spaces constructed through statistical depth measures. The effectiveness of this approach for detecting distributional changes was evaluated in different empirical studies by measuring the average detection delay and the proportion of false alarms. It is expected that this research can contribute to the application of this method in the online monitoring of machine learning models in dynamic environments, allowing the identification of the appropriate time to recalibrate or retrain algorithms in a more reliable manner.

Keywords: Dataset shift, Dataset Distribution change detection, Change point detection, Sequential data, Batch processing, Machine learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráficos da função g_i dada em (2.15) para os valores de $a = 0,3, 0,5, 0,9$ e 1 .	36
Figura 2 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de Poisson(2) utilizando o <i>transducer</i> determinístico , com α^1 (esquerda) e α^2 (direita).	41
Figura 3 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de Poisson(2) utilizando o <i>transducer</i> randomizado , com α^1 (esquerda) e α^2 (direita).	41
Figura 4 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de Normal(0,1) utilizando o <i>transducer</i> determinístico , com α^1 (esquerda) e α^2 (direita).	42
Figura 5 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de Normal(0,1) utilizando o <i>transducer</i> randomizado , com α^1 (esquerda) e α^2 (direita).	42
Figura 6 – Proporção de falsos alarmes (erro tipo I) para as diferentes distribuições do cenário descrito na tabela 5, em que à esquerda de cada par de gráficos, temos a medida de não conformidade α^1 e à direita, α^2 .	44
Figura 7 – Martingales do Experimento 1 (Normal), à esquerda, gerados por α^1 e à direita, por α^2 .	45
Figura 8 – Martingales do Experimento 2 (Gamma), à esquerda, gerados por α^1 e à direita, por α^2 .	46
Figura 9 – Martingales do Experimento 3 (Exponencial), à esquerda, gerados por α^1 e à direita, por α^2 .	46
Figura 10 – Martingales do Experimento 4 (Poisson), à esquerda, gerados por α^1 e à direita, por α^2 .	46
Figura 11 – Comparação dos métodos para distribuição Normal, com diferentes valores para μ_1 .	55
Figura 12 – AUC para distribuição Normal, com diferentes valores para μ_1 .	56
Figura 13 – Comparação dos métodos para distribuição Poisson, com diferentes valores para λ_1 .	57
Figura 14 – AUC para distribuição Poisson, com diferentes valores para λ_1 .	58
Figura 15 – Comparação dos métodos para distribuição Exponencial, com diferentes valores para λ_1 .	59
Figura 16 – AUC para distribuição Exponencial, com diferentes valores para λ_1 .	60

Figura 17 – Comparação dos métodos para distribuição Normal bivariada, com diferentes valores para o vetor de médias μ_1 , e matrizes de covariância $\Sigma_0 = \Sigma_1$ idênticas.	62
Figura 18 – AUC para distribuição Normal bivariada, com diferentes valores para o vetor de médias μ_1 , e matrizes de covariância idênticas $\Sigma_0 = \Sigma_1$.	63
Figura 19 – Comparação dos métodos para distribuição Normal bivariada, com diferentes valores para Σ_1 e ($\mu_0 = \mu_1$ idênticos).	64
Figura 20 – AUC para Normal Bivariada com diferentes valores para Σ_1 e ($\mu_0 = \mu_1$ idênticos).	64
Figura 21 – Comparação dos métodos para distribuição Normal Multivariada com $d = 10$, à esquerda, Configuração 1 da Tabela 8, à direita, Configuração 2.	66
Figura 22 – AUC para distribuição Normal Multivariada com $d = 10$, à esquerda, Configuração 1 da Tabela 8, à direita, Configuração 2.	66
Figura 23 – Resultados do experimento adotando valores errados para o caso univariado , à esquerda, as curvas da relação de taxa de alarme falsos por tempo de detecção e à direita, a área abaixo da curva.	69
Figura 24 – Resultados do experimento adotando valores errados para o caso multivariado , à esquerda, as curvas da relação de taxa de alarme falsos por tempo de detecção e à direita, a área abaixo da curva.	69
Figura 25 – Comparação entre as medidas de não conformidade e <i>transducers</i> para diferentes tipos de <i>shift</i> no Experimento.	77
Figura 26 – AUCs obtidas para diferentes tipos de <i>shift</i> no Experimento.	78

LISTA DE TABELAS

Tabela 1	– Cenários para análise dos valores p^c sob H_0	40
Tabela 2	– Percentual de réplicas nas quais a hipótese nula de uniformidade foi rejeitada no teste KS com $\alpha = 5\%$	40
Tabela 3	– Percentual de réplicas nas quais a hipótese nula de falta de autocorrelação foi rejeitada pelo teste de Durbin-Watson com $\alpha = 5\%$	43
Tabela 4	– Percentual de réplicas nas quais a hipótese nula de falta de autocorrelação foi rejeitada pelo teste de Box-Pierce-Ljung-Box com $\alpha = 5\%$	43
Tabela 5	– Cenários para análise dos Martingales	43
Tabela 6	– Distribuições utilizadas nos experimentos	53
Tabela 7	– Configurações do experimento multivariado $d = 2$	61
Tabela 8	– Configurações do experimento multivariado $d = 10$	64

LISTA DE ABREVIATURAS E SIGLAS

<i>CUSUM</i>	<i>Cumulative Sum Control Chart</i>
<i>KL</i>	<i>Kullback-Leibler</i>
<i>NN</i>	<i>Nearest Neighbour</i>
<i>PP</i>	<i>Posterior Probability</i>
<i>SR</i>	<i>Shirayev-Roberts</i>
<i>KS</i>	<i>Kolmogorov–Smirnov</i>

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Introdução	23
1.2	Estrutura do trabalho	25
2	MÉTODO <i>CONFORMAL</i> PARA TESTES SEQUENCIAIS DE PER- MUTABILIDADE	27
2.1	Permutabilidade e Aleatoriedade	27
2.2	Medidas de Não Conformidade e <i>Conformal Transducers</i>	28
2.3	<i>Conformal</i> Martingales e Supermartingales	31
2.3.1	<i>Martingale de Potência (Power Martingale)</i>	35
2.3.2	<i>Fluxograma resumo</i>	37
3	ANÁLISE EMPÍRICA DA VALIDADE DO MÉTODO <i>CONFORMAL</i>	39
3.1	Análise dos p^c	39
3.2	Análise dos Martingales.	43
3.2.1	<i>Controle do Erro do Tipo I</i>	44
3.2.2	<i>Análise dos valores dos martingales.</i>	45
4	COMPARAÇÃO EMPÍRICA ENTRE MÉTODOS CLÁSSICOS E MÉTODOS BASEADOS EM <i>CONFORMAL</i> MARTINGALES	49
4.1	Métodos clássicos	50
4.1.1	<i>CUSUM (Cumulative Sum):</i>	50
4.1.2	<i>Shiryayev-Roberts (SR):</i>	51
4.1.3	<i>Probabilidade Posterior (PP):</i>	52
4.2	Experimentos Univariados	52
4.3	Experimentos Multivariados	58
4.3.1	<i>Experimentos Multivariados de dimensão $d = 2$</i>	60
4.3.2	<i>Experimentos Multivariados de dimensão $d = 10$</i>	62
4.4	Especificação Incorreta em Métodos Clássicos	67
5	ESTUDOS EM LOTES DE DADOS SEQUENCIAIS	71
5.1	Definição do problema	72
5.2	<i>Kullback-Leibler</i>	73
5.2.1	<i>Estimativa da divergência de Kullback-Leibler</i>	74

5.3	Definição dos experimentos	75
5.4	Simulação	76
5.4.1	<i>Experimento</i>	76
6	CONCLUSÕES E CONSIDERAÇÕES FINAIS	81
	REFERÊNCIAS	83
APÊNDICE A	DEDUÇÕES DAS FORMAS RECURSIVAS DOS MÉ- TODOS	87
A.1	Método PP (Probabilidade Posterior)	87
A.2	Método CUSUM	88
A.3	Método Shiryaev–Roberts	89

INTRODUÇÃO

1.1 Introdução

A análise de dados é um campo de estudo em constante evolução, impulsionado pela crescente disponibilidade de informações e pela diversificação dos campos de aplicação. No entanto, essa diversificação também traz desafios significativos, como as mudanças na distribuição dos dados ao longo do tempo. Em muitas aplicações, especialmente na área de aprendizado de máquina, assume-se que os dados de treinamento são provenientes de observações independentes e identicamente distribuídas (*i.i.d*). No entanto, essa suposição pode não se manter válida ao longo do tempo, o que compromete a confiabilidade dos modelos e das análises baseadas nesses dados.

Mudanças na distribuição podem ocorrer de diversas formas e impactar diretamente a precisão das previsões e decisões tomadas a partir dos modelos estatísticos e de aprendizado de máquina. Se o mecanismo gerador dos dados sofrer alterações, as previsões dos modelos treinados sob a hipótese de *i.i.d* podem se tornar inconsistentes e pouco confiáveis. Nesse contexto, torna-se essencial o desenvolvimento de técnicas estatísticas capazes de detectar rapidamente tais mudanças para permitir a adaptação e o retreinamento dos modelos. Um conjunto de desafios particularmente relevante nesse cenário é conhecido como *dataset shift*.

A identificação de mudanças na distribuição dos dados pode ser crucial em diversos contextos. Por exemplo, em uma empresa que realiza previsões de demanda, os dados de vendas são atualizados mensalmente, e qualquer alteração na distribuição desses dados pode impactar diretamente as estratégias de negócios. De maneira mais geral, pode-se desejar analisar uma sequência de observações para identificar se, a partir de um determinado ponto, os dados passaram a ser gerados por uma distribuição de probabilidade diferente, algo comumente abordado na literatura sob o conceito de *detecção de ponto de mudança*.

No contexto da modelagem preditiva, os *dataset shifts* podem assumir diferentes formas,

como *covariate shift*, *total dataset shift*, *label shift* e *concept shift* (MORENO-TORRES *et al.*, 2012b). O *covariate shift* ocorre quando a distribuição das características (covariáveis) dos dados muda ao longo do tempo. Já o *total dataset shift* envolve uma mudança completa na distribuição dos dados, afetando tanto as covariáveis quanto a variável resposta. O *label shift* refere-se a alterações na distribuição das variáveis resposta, enquanto o *concept shift* indica mudanças nos conceitos subjacentes, impactando as distribuições condicionais dos dados. Esses fenômenos podem comprometer significativamente o desempenho dos modelos preditivos e, portanto, a detecção dessas mudanças é essencial para a manutenção da qualidade das previsões.

Neste trabalho estudamos um método que pode ser utilizado para detectar desde mudanças na distribuição de sequências de variáveis aleatórias univariadas e multivariadas, até identificar e quantificar diferentes tipos de *dataset shift* em sequências de conjuntos de dados. O método desenvolvido baseia-se na teoria de *conformal prediction* (VOVK; NOURETDINOV; GAMMERMAN, 2003), em que os autores constroem uma estatística sequencial baseada na sequência acumulada de observações, que é um martingale conformal (*CM - conformal martingale*), sob a hipótese nula de que os dados são permutáveis. A partir disso, ele propõe uma regra de detecção que controla o erro do tipo I (alarme falso).

Vovk propôs originalmente uma metodologia para detectar mudanças em sequências de observações dentro de um mesmo conjunto de dados, utilizando distâncias usuais, como a euclidiana, para medir dissimilaridades. Neste trabalho, além de validar empiricamente essa abordagem para diferentes distribuições, exploramos diferentes medidas de não conformidade e comparamos seu desempenho com métodos clássicos de detecção de mudança sob diferentes cenários de geração dos dados, considerando tanto o caso univariado quanto multivariado.

As principais contribuições dessa adaptação residem no uso da medida *Kullback-Leibler* (*KL*) (KULLBACK; LEIBLER, 1951), permitindo em um contexto preditivo, a detecção de mudanças em lotes sequenciais, nos quais múltiplos conjuntos de dados chegam em uma sequência contínua. Além disso, para o contexto de distribuições multivariadas, foram incorporadas novas métricas baseadas em profundidade estatística (MOSLER, 2013), ampliando a aplicabilidade do método. A métrica *KL*, amplamente reconhecida na teoria da informação, quantifica diferenças entre distribuições de probabilidade e se mostra uma ferramenta poderosa para detectar variações na distribuição dos dados.

Além disso, utilizando *conformal prediction*, dispomos de um método de **detecção de ponto de mudança** que não requer especificação prévia das distribuições antes e depois da mudança, diferentemente de abordagens clássicas como *Cumulative Sum Control Chart* (*CUSUM*) (PAGE, 1954), *Shirayev-Roberts* (*SR*) (SHIRYAEV, 1961) e *Posterior Probability* (*PP*) (SHIRYAEV, 1963), que exigem esse conhecimento prévio.

Ao longo desta dissertação, exploramos a teoria subjacente aos métodos propostos, sua implementação e sua avaliação em diversos cenários. Foram realizadas simulações e estudos empíricos para validar a eficácia da abordagem, fornecendo evidências concretas de sua aplicabi-

lidade. Espera-se que esta pesquisa contribua para a detecção e compreensão de mudanças em dados sequenciais, beneficiando diversas áreas, como monitoramento de sistemas de aprendizado de máquina em produção, detecção de fraudes financeiras, análise de séries temporais em contextos econômicos e industriais e qualquer que seja o cenário em que é importante detectar se houve alguma mudança nos dados ao longo do tempo.

1.2 Estrutura do trabalho

Este trabalho está organizado da seguinte maneira:

- No Capítulo 2 é realizada a introdução ao método *conformal*, apresentando as principais estruturas que servem de base para este trabalho.
- No Capítulo 3 são apresentadas validações empíricas sobre a validade do método *conformal* através de simulações.
- No Capítulo 4 é realizada uma primeira comparação dos métodos propostos por esse trabalho com métodos clássicos utilizados para detecção de ponto de mudança (casos univariados e multivariados).
- No Capítulo 5 a análise de ponto de mudança é expandida para o contexto de preditivo, em que são feitos testes para detecção de diferentes tipos de *shifts* em conjuntos de dados sequenciais.
- No Capítulo 6 são apresentadas as considerações finais do trabalho.

MÉTODO *CONFORMAL* PARA TESTES SEQUENCIAIS DE PERMUTABILIDADE

Neste capítulo apresentamos brevemente a teoria do método conformal proposto em [Vovk, Nouretdinov e Gammerman \(2003\)](#) para detecção de mudança de distribuição em dados sequenciais, a qual serve de base para os métodos estudados neste trabalho. Começamos abordando conceitos cruciais, tais como os conceitos de permutabilidade e aleatoriedade e, em seguida, discutimos os principais aspectos do método conformal: a construção de *transducers* e martingales de teste para a hipótese de permutabilidade a partir da consideração de medidas de não conformidade.

2.1 Permutabilidade e Aleatoriedade

Considere que o vetor de observações aleatórias $\mathbf{Z} = (Z_1, Z_2, \dots)$ assume valores em um conjunto $\mathcal{Z}^\infty = \{(z_1, z_2, \dots) : z_i \in \mathcal{Z}, i \geq 1\}$ com distribuição de probabilidades Q . A *hipótese de aleatoriedade* afirma que a distribuição de probabilidades $Q = \mathbb{P}^\infty = \mathbb{P} \times \mathbb{P} \cdots$, isto é, as componentes Z_1, Z_2, \dots de \mathbf{Z} são independentes e identicamente distribuídas (i.i.d) com distribuição \mathbb{P} sobre \mathcal{Z} . Por outro lado, a *hipótese de permutabilidade* afirma que a distribuição Q é permutável, isto é, para todo $n \geq 1$ e toda permutação $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$

$$(Z_1, \dots, Z_n) \stackrel{\mathcal{D}}{=} (Z_{\pi(1)}, \dots, Z_{\pi(n)}),$$

cujo o símbolo $\stackrel{\mathcal{D}}{=}$ significa igualdade de distribuição.

A propriedade de permutabilidade é menos restritiva que a de independência, pois toda sequência que é i.i.d também é permutável ([SCHERVISH, 1995](#), p. 75). A propriedade de permutabilidade será fundamental para a validade do método conformal proposto por [Vovk, Nouretdinov e Gammerman \(2003\)](#).

2.2 Medidas de Não Conformidade e *Conformal Transducers*

No cerne do método conformal está a sequência de **medidas de não conformidade** $\{A_n : n \geq 1\}$, em que $A_n : \mathcal{Z}^n \rightarrow \mathbb{R}^n$, $n \geq 1$, associa a cada sequência $(z_1, z_2, \dots, z_n) \in \mathcal{Z}^n$ uma sequência $(\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{R}^n$, em que cada α_i representa o quão atípica é a observação z_i com respeito ao vetor (z_1, z_2, \dots, z_n) . Cada função A_n satisfaz a suposição de permutabilidade, isto é, para todo $n \geq 1$ e toda permutação $\pi : \{1, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ devemos ter

$$A_n(z_1, \dots, z_n) = (\alpha_1, \dots, \alpha_n) \implies A_n(z_{\pi(1)}, \dots, z_{\pi(n)}) = (\alpha_{\pi(1)}, \dots, \alpha_{\pi(n)}). \quad (2.1)$$

Uma forma mais simples de representar cada função A_n é defini-la em termos de uma função de discrepância $\rho_n : \mathcal{Z} \times \mathcal{Z}^n \rightarrow [0, +\infty]$, que quantifica a estranheza (ou surpresa) de $z \in \mathcal{Z}$ em relação a $(z_1, \dots, z_n) \in \mathcal{Z}^n$ que satisfaça

$$\rho_n(z, (z_1, \dots, z_n)) = \rho_n(z, (z_{\pi(1)}, \dots, z_{\pi(n)})),$$

para todo $z \in \mathcal{Z}$, todo $n \geq 1$ e toda permutação $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$.

A partir disto, definimos A_n como

$$A_n(z_1, \dots, z_n) = (\rho_n(z_1, (z_1, \dots, z_n)), \dots, \rho_n(z_n, (z_1, \dots, z_n))), \quad (2.2)$$

em que $(z_1, \dots, z_n) \in \mathcal{Z}^n$, $n \geq 1$. A representação de A_n dada em (2.2) satisfaz a condição de permutabilidade (2.1).

A partir da sequência de medidas de não conformidade $\{A_n : n \geq 1\}$, definimos o *conformal transducer* que converte estas medidas de não conformidade em uma sequência de valores em $[0, 1]$, que são interpretados como valores- p associados à hipótese de permutabilidade. Consideramos dois métodos para a construção de um conformal transducer: o método *determinístico* e o método *randomizado*.

O **conformal transducer determinístico** é uma sequência de funções $\{f_n : n \geq 1\}$, $f_n : \mathcal{Z}^n \rightarrow [0, 1]$, $n \geq 1$, definida como

$$f_n(z^n) := \frac{\#\{i \in \{1, \dots, n\} : \alpha_i^n \geq \alpha_n^n\}}{n}, \quad (2.3)$$

em que $z^n = (z_1, \dots, z_n) \in \mathcal{Z}^n$, $A_n(z_1, \dots, z_n) = (\alpha_1^n, \dots, \alpha_n^n)$ e o símbolo $\#(A)$ denota o número de elementos de um conjunto A . Na etapa $n \geq 1$, temos as observações z_1, \dots, z_n e $f_n(z_1, \dots, z_n)$ é a proporção de observações z_i com medidas de não conformidade mais extremas que a observação z_n , consequentemente valores pequenos de $f_n(z_1, \dots, z_n)$ indicam uma alta não conformidade de z_n com respeito às observações z_1, \dots, z_n . Desta forma, para cada $n \geq 1$, definimos o **valor- p conformal determinístico** como

$$p_n^c(\mathbf{z}) = f_n(z^n), \quad (2.4)$$

em que $z^n = (z_1, \dots, z_n)$ e $\mathbf{z} = (z_1, z_2, \dots) \in \mathcal{Z}^\infty$.

A partir da Expressão (2.3), podemos notar que, para cada $\mathbf{z} \in \mathcal{Z}^\infty$ fixado, $p_n^c(\mathbf{z})$ assume valores no conjunto finito $\{1/n, 2/n, \dots, (n-1)/n, 1\}$ e, caso haja empates com o valor α_n^n , este conjunto será ainda menor. A seguir, nesta seção, vemos que esta característica discreta dos valores-p conformais determinísticos está associada a uma perda de poder do teste.

Para contornar esse problema, consideramos o **conformal transducer randomizado**. Primeiramente, considere $\theta_1, \theta_2, \dots$ uma sequência de variáveis i.i.d com distribuição uniforme no intervalo $(0, 1)$ em que $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ é independente de $\mathbf{Z} = (Z_1, Z_2, \dots)$, isto é, os valores de $\boldsymbol{\theta}$ são ruídos independentes das observações. Definimos o *transducer* randomizado como a sequência $\{\tilde{f}_n : n \geq 1\}$, $\tilde{f}_n : \mathcal{Z}^n \times (0, 1)^n \rightarrow [0, 1]$, $n \geq 1$, como

$$\tilde{f}_n(z^n, \boldsymbol{\theta}^n) := \frac{\#\{i \in \{1, \dots, n\} : \alpha_i^n > \alpha_n^n\} + \theta_n \#\{i \in \{1, \dots, n\} : \alpha_i^n = \alpha_n^n\}}{n}, \quad (2.5)$$

em que $z^n = (z_1, \dots, z_n) \in \mathcal{Z}^n$, $\boldsymbol{\theta}^n = (\theta_1, \dots, \theta_n) \in [0, 1]^n$ e $A_n(z_1, \dots, z_n) = (\alpha_1^n, \dots, \alpha_n^n)$. A interpretação de \tilde{f}_n é a mesma que a de f_n e, similarmente, definimos para cada $n \geq 1$ o **valor-p conformal randomizado** como

$$\tilde{p}_n^c(\mathbf{z}, \boldsymbol{\theta}) = \tilde{f}_n(z^n, \boldsymbol{\theta}^n), \quad (2.6)$$

em que $z^n = (z_1, \dots, z_n) \in \mathcal{Z}^n$, $\mathbf{z} = (z_1, z_2, \dots) \in \mathcal{Z}^\infty$, $\boldsymbol{\theta}^n = (\theta_1, \dots, \theta_n) \in [0, 1]^n$ e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in (0, 1)^\infty$. Observe que agora os valores-p randomizados gerados não dependem exclusivamente do vetor de observações \mathbf{z} . No entanto, a partir das expressões (2.3) e (2.5), temos que

$$\tilde{f}_n(z^n, \boldsymbol{\theta}^n) \leq f_n(z^n), \quad \text{em que } z^n = (z_1, \dots, z_n), \boldsymbol{\theta}^n = (\theta_1, \dots, \theta_n), \quad (2.7)$$

para todo $n \geq 1$, $\mathbf{z} = (z_1, z_2, \dots) \in \mathcal{Z}^\infty$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in [0, 1]^\infty$.

Como $\theta_n \leq 1$ para todo $n \geq 1$, segue de (2.7) que

$$\tilde{p}_n^c(\mathbf{z}, \boldsymbol{\theta}) \leq p_n^c(\mathbf{z}) \quad (2.8)$$

para todo $\mathbf{z} = (z_1, z_2, \dots) \in \mathcal{Z}^\infty$ e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in (0, 1)^\infty$.

Vale destacar que o *transducer* randomizado, conforme definido em (2.5), é um *E/U (Exchangeability-to-Uniformity) transducer*, ou seja, sob a hipótese de permutabilidade, ele gera valores-p i.i.d com distribuição uniforme no intervalo $(0, 1)$. Este resultado, que é fundamental para a construção de martingales que controlam precisamente o erro do tipo I em testes sequenciais, é apresentado a seguir e sua demonstração pode ser encontrada em [Vovk, Nouretdinov e Gammerman \(2003\)](#).

Teorema 1. Se $\mathbf{Z} = (Z_1, Z_2, \dots)$ tem distribuição Q , Q é uma distribuição de probabilidade permutável e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ tem distribuição i.i.d uniforme em $(0, 1)$ independente de \mathbf{Z} , então

$$\tilde{p}_1^c(\mathbf{Z}, \boldsymbol{\theta}), \tilde{p}_2^c(\mathbf{Z}, \boldsymbol{\theta}), \dots \stackrel{i.i.d}{\sim} U(0, 1).$$

A partir do resultado do Teorema 1, concluímos que a probabilidade de $[\tilde{p}_n^c(\mathbf{Z}, \boldsymbol{\theta}) \leq u]$ é exatamente igual a u para todo $u \in [0, 1]$ e todo $n \geq 1$. Por outro lado, o *transducer* determinístico, como definido em (2.3), não produz valores-p conformais com distribuição uniforme exata. Em vez disso, os valores gerados tem distribuição **super-uniforme** (BENJAMINI; YEKUTIELI, 2001): uma variável X tem distribuição super-uniforme caso $P(X \leq u) \leq u$ para todo $u \in [0, 1]$. Por essa razão, o *transducer* determinístico é denominado *E/U supertransducer*. Este resultado é apresentado no corolário a seguir e é consequência imediata do Teorema 1 e da inequação (2.8).

Corolário 1. *Se $\mathbf{Z} = (Z_1, Z_2, \dots)$ tem distribuição Q , em que Q é uma distribuição de probabilidade permutável, então a sequência de variáveis aleatórias $\{p_n^c(\mathbf{Z}) : n \geq 1\}$ satisfaz*

$$Q([\tilde{p}_n^c(\mathbf{Z}) \leq u]) \leq u, \quad \text{para todo } u \in [0, 1] \text{ e todo } n \geq 1.$$

Demonstração. Primeiramente, considere $\mathbf{Z} = (Z_1, Z_2, \dots)$ com distribuição Q sobre \mathcal{Z}^∞ e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ com distribuição U^∞ sobre $(0, 1)^\infty$, em que \mathbf{Z} e $\boldsymbol{\theta}$ são independentes. Combinando \mathbf{Z} e $\boldsymbol{\theta}$ em uma sequência $\tilde{\mathbf{Z}} = (Z_1, \theta_1, Z_2, \theta_2, \dots)$ com distribuição \tilde{Q} sobre $(\mathcal{Z} \times (0, 1))^\infty$, que pode ser determinada a partir das distribuições finito-dimensionais \tilde{Q}_n definidas como

$$\tilde{Q}_n(Z_1 \leq z_1, \theta_1 \leq u_1, \dots, Z_n \leq z_n, \theta_n \leq u_n) = Q(Z_1 \leq z_1, \dots, Z_n \leq z_n) \times U^\infty(\theta_1 \leq u_1, \dots, \theta_n \leq u_n) \quad (2.9)$$

para quaisquer $z_1, \dots, z_n \in \mathcal{Z}$, $u_1, \dots, u_n \in \mathbb{R}$ e todo $n \geq 1$. A sequência de distribuições finito-dimensionais dada em (2.9) sobre o espaço $(\mathcal{Z} \times (0, 1))^n$ satisfaz a condição de compatibilidade do Teorema de Extensão de Kolmogorov para sequências aleatórias (Teorema 6.3.4, p. 207, (ATHREYA; LAHIRI, 2006)). Observe que a distribuição Q pode ser obtida marginalizando \tilde{Q} com respeito às componentes com índice par (associadas ao vetor $\boldsymbol{\theta}$).

Segue da inequação (2.8) que

$$p_n^c(\mathbf{Z}) \geq \tilde{p}_n^c(\mathbf{Z}, \boldsymbol{\theta}), \quad \text{para todo } n \geq 1,$$

e do Teorema 1 sabemos que $\tilde{p}_1^c(\mathbf{Z}, \boldsymbol{\theta})$, $\tilde{p}_2^c(\mathbf{Z}, \boldsymbol{\theta})$, \dots são i.i.d com distribuição $U(0, 1)$. Portanto, para todo $n \geq 1$ e todo $u \in [0, 1]$, temos que

$$Q([\tilde{p}_n^c(\mathbf{Z}) \leq u]) = \tilde{Q}([\tilde{p}_n^c(\mathbf{Z}) \leq u]) \leq \tilde{Q}([\tilde{p}_n^c(\mathbf{Z}, \boldsymbol{\theta}) \leq u]) \leq u.$$

□

A partir da inequação (2.8) concluímos que os valores-p conformais determinísticos constituem uma estratégia mais conservadora para testar a hipótese de permutabilidade, mas que ainda mantém o controle do erro do tipo I de acordo com o Corolário 1.

Em Vovk, Nouretdinov e Gammernan (2003), os autores ilustram o método em um problema de classificação em que $z_n = (x_n, y_n)$, com x_n sendo as covariáveis (*features*) e y_n uma variável resposta categórica (*label*). Mais especificamente, eles consideram o conjunto de dados

USPS (descrito em [Vapnik \(1998\)](#)), em que uma observação (x, y) é constituída de uma imagem x (matriz 16×16 de pixels) e $y \in \{0, 1, \dots, 9\}$ é um rótulo associado à imagem. A medida de discrepância que foi adotada naquele trabalho, baseada no método de classificação *Nearest Neighbour* (NN), é dada por

$$\rho_n((x, y), (z_1, \dots, z_n)) = \frac{\min\{d(x, x_i) : y_i = y, x \neq x_i\}}{\min\{d(x, x_i) : y_i \neq y\}},$$

isto é, a não conformidade de uma observação $z = (x, y)$ em relação ao conjunto de observações z_1, \dots, z_n aumenta caso a distância mínima para observações com rótulos distintos diminua (denominador) e diminui caso a distância mínima para observações com o mesmo rótulo diminua. A observação (x, y) é considerada estranha ou não conforme caso tenha imagens semelhantes a observações com rótulos distintos ou tenha imagens diferentes de observações com mesmo rótulo. Neste trabalho, foram exploradas outras alternativas além da discrepância baseada em NN.

2.3 Conformal Martingales e Supermartingales

Nesta seção, denotamos a sequência dos valores-p conformais determinísticos por $\{p_n^c : n \geq 1\}$ e a sequência dos valores-p randomizados por $\{\tilde{p}_n^c : n \geq 1\}$, omitindo os argumentos \mathbf{z} e θ , que ficam subentendidos. A partir da sequência de valores-p conformais, construímos as sequências $\{S_n : n \geq 1\}$ e $\{\tilde{S}_n : n \geq 1\}$ como

$$S_n = \prod_{i=1}^n g_i(p_i^c), \quad n \geq 1 \quad (2.10)$$

$$\tilde{S}_n = \prod_{i=1}^n g_i(\tilde{p}_i^c), \quad n \geq 1, \quad (2.11)$$

em que $g_i : (0, 1] \rightarrow [0, +\infty)$, $i \geq 1$, é uma função que satisfaz as seguintes condições para todo $i \geq 1$:

$$(a) \quad E[g_i(U)] = 1 \text{ para } U \sim U(0, 1); \quad (2.12)$$

$$(b) \quad g_i(x_1) \geq g_i(x_2) \text{ para quaisquer } x_1 < x_2. \quad (2.13)$$

A função g_i é chamada de *betting function* e assume valores grandes quando o i -ésimo valor-p conformal é pequeno, $i \geq 1$. A seguir, construímos uma sequência $\{\tilde{S}_n : n \geq 0\}$, $\tilde{S}_0 = 1$ a partir das funções g_n e dos valores-p conformais randomizados \tilde{p}_n^c em que \tilde{S}_n pode ser interpretado como capital acumulado até a etapa n por um jogador que aposta contra a hipótese de permutabilidade. Caso o jogador acumule capital apostando contra esta hipótese, isso coloca dúvidas sobre a validade da mesma. O jogo pode ser explicado em etapas da seguinte forma:

- Na etapa inicial ($i = 0$), o jogador possui uma fortuna $\tilde{S}_0 = 1$;

- Ao final da etapa 1, a fortuna atual do jogador ($\tilde{S}_0 = 1$) será multiplicada por um fator $g_1(\tilde{p}_1^c)$, fazendo com que sua fortuna seja de $\tilde{S}_1 = g_1(\tilde{p}_1^c)$;
- Ao final da etapa 2, a fortuna atual do jogador ($\tilde{S}_1 = g_1(\tilde{p}_1^c)$) será multiplicada por um fator $g_2(\tilde{p}_2^c)$, fazendo com que sua fortuna seja de $\tilde{S}_2 = g_1(\tilde{p}_1^c) g_2(\tilde{p}_2^c)$;
- Ao final da etapa n , a fortuna atual do jogador (\tilde{S}_{n-1}) será multiplicada por um fator $g_n(\tilde{p}_n^c)$, fazendo com que sua fortuna seja de $\tilde{S}_n = \tilde{S}_{n-1} g_n(\tilde{p}_n^c)$, $n > 1$.

Observe que, de acordo com a suposição (b) dada em (2.13), o jogador é mais favorecido quanto menor for o valor-p gerado (evidência contrária a hipótese nula). Além disso, de acordo com a suposição (a) dada em (2.12), o fator de ganho em média a cada etapa é 1. O próximo resultado garante que a sequência $\{\tilde{S}_n : n \geq 0\}$, $\tilde{S}_0 = 1$, é um *exchangeability martingale*, isto é, é um martingale sempre que a distribuição Q de $\mathbf{Z} = (Z_1, Z_2, \dots)$ for permutável.

Teorema 2. Para $\mathbf{Z} = (Z_1, Z_2, \dots)$ com distribuição Q permutável, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ com distribuição U^∞ sobre $(0, 1)^\infty$, definimos $\tilde{\mathbf{Z}} = (Z_1, \theta_1, Z_2, \theta_2, \dots)$ com distribuição \tilde{Q} sobre $(\mathcal{Z} \times (0, 1))^\infty$ (de acordo com (2.9)), a sequência $\{\tilde{p}_n^c(\mathbf{Z}, \boldsymbol{\theta}) : n \geq 1\}$ como em (2.8) e (2.5), a sequência $\{\tilde{S}_n : n \geq 1\}$ como em (2.11) e a filtração $\mathcal{F} = \{\mathcal{F}_n : n \geq 0\}$ em que $\mathcal{F}_0 = \{\emptyset, (\mathcal{Z} \times (0, 1))^\infty\}$, $\mathcal{F}_n = \sigma(\tilde{p}_n^c)$, $n \geq 1$. Sob estas hipóteses, temos que a sequência $\{\tilde{S}_n : n \geq 0\}$, em que $\tilde{S}_0 = 1$, é um martingale não negativo com respeito à filtração \mathcal{F} , isto é,

- (i) \tilde{S}_n é mensurável com respeito a \mathcal{F}_n para todo $n \geq 0$ (a sequência é adaptada à filtração);
- (ii) $\tilde{S}_n \geq 0$ para todo $n \geq 0$;
- (iii) $E^{\tilde{Q}}[\tilde{S}_n] < +\infty$ para todo $n \geq 0$;
- (iv) $E^{\tilde{Q}}[\tilde{S}_{n+1} | \mathcal{F}_n] = \tilde{S}_n$ para todo $n \geq 0$.

Demonstração. Para mostrar (i), note que \tilde{S}_n é função dos n primeiros valores-p conformais randomizados $\tilde{p}_1^c, \dots, \tilde{p}_n^c$ e, portanto, é mensurável com respeito a \mathcal{F}_n para todo $n \geq 1$. Como S_0 é constante, é mensurável com respeito a \mathcal{F}_0 .

A condição (ii) segue do fato das *betting functions* assumirem valores em $[0, +\infty)$.

A condição (iii) segue do fato da sequência $\{\tilde{p}_n^c : n \geq 1\}$ ser i.i.d $U(0, 1)$ (Teorema 1) e da propriedade (a) das *betting functions* dada em (2.12). De fato,

$$E^{\tilde{Q}}[\tilde{S}_n] = E^{\tilde{Q}} \left[\prod_{i=1}^n g_i(\tilde{p}_i^c) \right] = \prod_{i=1}^n E^{\tilde{Q}} [g_i(\tilde{p}_i^c)] = 1 < +\infty.$$

Em seguida, para mostrar (iv) observe que

$$\begin{aligned}
E^{\tilde{Q}}[\tilde{S}_{n+1} \mid \mathcal{F}_n] &= E^{\tilde{Q}}\left[\prod_{i=1}^{n+1} g_i(\tilde{p}_i^c) \mid \mathcal{F}_n\right] \\
&= E^{\tilde{Q}}\left[g_{n+1}(\tilde{p}_{n+1}^c) \prod_{i=1}^n g_i(\tilde{p}_i^c) \mid \mathcal{F}_n\right] \\
&= \left(\prod_{i=1}^n g_i(\tilde{p}_i)\right) E^Q[g_{n+1}(\tilde{p}_{n+1}^c) \mid \mathcal{F}_n] \\
&= \left(\prod_{i=1}^n g_i(\tilde{p}_i)\right) E^Q[g_{n+1}(\tilde{p}_{n+1}^c)] \\
&= \left(\prod_{i=1}^n g_i(\tilde{p}_i)\right) \\
&= \tilde{S}_n
\end{aligned}$$

em que na antepenúltima passagem usamos a independência entre \tilde{p}_{n+1}^c e a sequência $(\tilde{p}_1^c, \dots, \tilde{p}_n^c)$, na penúltima a suposição (a) dada em (2.12) para as *betting function* e o fato de $\tilde{p}_{n+1}^c \sim U(0, 1)$ (Teorema 1), e na última passagem a definição de \tilde{S}_n dada em (2.11). \square

A partir deste resultado, sabemos que caso as observações tenham distribuição Q permutável, então o martingale $\{\tilde{S}_n : n \geq 0\}$ oscila em torno do seu valor inicial $S_0 = 1$. Portanto, uma evidência contrária a hipótese de permutabilidade é fornecida quando observamos valores grandes de \tilde{S}_n . Uma garantia formal para esta última interpretação é fornecida pela desigualdade de Ville (VILLE, 1939) que afirma que

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} [X_n \geq c]\right) \leq \frac{1}{C}, \quad \text{para todo } C > 0, \quad (2.14)$$

sempre que $\{X_n : n \geq 0\}$ for um supermartingale não negativo com valor inicial $X_0 = 1$ com respeito a uma filtração $\mathcal{F} = \{\mathcal{F}_n : n \geq 0\}$, isto é,

- X_n é mensurável com respeito a \mathcal{F}_n para todo $n \geq 0$;
- $X_n \geq 0$ para todo $n \geq 0$;
- $E[X_n] < +\infty$ para todo $n \geq 0$;
- $E[X_{n+1} \mid \mathcal{F}_n] \leq X_n$ para todo $n \geq 0$.

Note que todo martingale é um supermartingale, então podemos concluir que para a sequência \tilde{S}_n definida em (2.11) vale

$$\tilde{Q}\left(\bigcup_{n=1}^{+\infty} [\tilde{S}_n \geq C]\right) \leq \frac{1}{C}, \quad \text{para todo } C > 0$$

O resultado a seguir caracteriza a sequência $\{S_n : n \geq 1\}$ definida em (2.10) como um *exchangeability supermartingale*, isto é, é um supermartingale sempre que a distribuição Q de $\mathbf{Z} = (Z_1, Z_2, \dots)$ for permutável.

Teorema 3. Para $\mathbf{Z} = (Z_1, Z_2, \dots)$ com distribuição Q permutável, definimos a sequência $\{p_n^c(\mathbf{Z}) : n \geq 1\}$ como em (2.4) e (2.3), a sequência $\{S_n : n \geq 1\}$ como em (2.10) e a filtração $\mathcal{F} = \{\mathcal{F}_n : n \geq 0\}$ em que $\mathcal{F}_0 = \{\emptyset, \mathcal{Z}^\infty\}$, $\mathcal{F}_n = \sigma(p_n^c)$, $n \geq 1$. Sob estas hipóteses, temos que a sequência $\{S_n : n \geq 0\}$, em que $S_0 = 1$, é um supermartingale não negativo com respeito à filtração \mathcal{F} , isto é,

- (i) S_n é mensurável com respeito a \mathcal{F}_n para todo $n \geq 0$ (a sequência é adaptada à filtração);
- (ii) $S_n \geq 0$ para todo $n \geq 0$;
- (iii) $E^Q[S_n] < +\infty$ para todo $n \geq 0$;
- (iv) $E^Q[S_{n+1} | \mathcal{F}_n] \leq S_n$ para todo $n \geq 0$.

Demonstração. Para mostrar (i), note que S_n é função dos n primeiros valores- p conformais determinísticos p_1^c, \dots, p_n^c e, portanto, é mensurável com respeito a \mathcal{F}_n para todo $n \geq 1$. Como S_0 é constante, é mensurável com respeito a \mathcal{F}_0 .

A condição (ii) segue do fato das *betting functions* assumirem valores em $[0, +\infty)$.

Para verificar a condição (iii), note que $p_n^c \geq \tilde{p}_n^c$ para todo $n \geq 1$ (desigualdade (2.7)) e, da propriedade (b) das *betting functions* dada em (2.13), segue que

$$g_n(p_n^c) \leq g_n(\tilde{p}_n^c), \quad \text{para todo } n \geq 1.$$

Disto e do fato da sequência $\{\tilde{p}_n^c : n \geq 1\}$ ser i.i.d $U(0, 1)$ (Teorema 1), segue que

$$E^Q[S_n] = E^{\tilde{Q}}[S_n] = E^{\tilde{Q}} \left[\prod_{i=1}^n g_i(p_i^c) \right] \leq E^{\tilde{Q}} \left[\prod_{i=1}^n g_i(\tilde{p}_i^c) \right] = 1 < +\infty.$$

Para mostrar (iv), note que segue de (2.8) que $p_n^c \geq \tilde{p}_n^c$ para todo $n \geq 1$, o que implica pela propriedade (b) das *betting function* dada em (2.13) que $g_n(p_n^c) \leq g_n(\tilde{p}_n^c)$ para todo $n \geq 1$. Portanto,

$$S_n = \prod_{i=1}^n g_i(p_i^c) \leq \prod_{i=1}^n g_i(\tilde{p}_i^c) = \tilde{S}_n, \quad \text{para todo } n \geq 1,$$

de onde segue que

$$E^Q[S_{n+1} | \mathcal{F}_n] \leq E^Q[\tilde{S}_{n+1} | \mathcal{F}_n] = S_n, \quad \text{para todo } n \geq 1,$$

na qual usamos o fato de $\{\tilde{S}_n\}$ ser um martingale dado no Teorema 2. □

Os Teoremas 2 e 3 caracterizam as sequências $\{S_n : n \geq 1\}$ e $\{\tilde{S}_n : n \geq 1\}$ definidas em (2.10) e (2.11) como um *exchangeability supermartingale* e um *exchangeability martingale*, respectivamente. De acordo com a desigualdade de Ville dada em (2.14), duas possíveis regras de decisão de nível α são dadas por rejeitar a hipótese nula de permutabilidade caso para algum $n \geq 1$ tenhamos $S_n \geq 1/\alpha$ ou rejeitar esta hipótese caso algum $\tilde{S}_n \geq 1/\alpha$. Portanto, o uso de ambas as regras com S_n ou \tilde{S}_n fornecem testes sequenciais válidos para a hipótese de permutabilidade. No entanto, a regra de decisão usando S_n é mais conservadora, uma vez que $S_n \leq \tilde{S}_n$ para todo $n \geq 1$.

Todo martingale é um supermartingale, pois satisfaz $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$, o que implica $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$;

Já o contrário não é verdadeiro: supermartingales permitem a expectativa condicional decrescer, mas não exigem igualdade.

2.3.1 Martingale de Potência (Power Martingale)

Uma possibilidade para a escolha das *betting functions* é dada por

$$g_i(p) = ap^{a-1}, \quad p \in (0, 1], \quad (2.15)$$

em que $i \geq 1$ e $a \in (0, 1]$ é um valor fixado. A função g_i , dada em 2.15, pertence à família de densidades Beta(a, b), cuja expressão genérica é dada por

$$h(p | a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}, \quad 0 < p < 1.$$

Portanto, g_i é uma densidade Beta com parâmetros a e $b = 1$, em que supomos $0 < a \leq 1$. A Figura 1 ilustra o gráfico da função g_i para diferentes casos do parâmetro a . Podemos observar nesta figura que o caso $a = 1$ corresponde a uma constante igual a 1 e para os demais casos temos uma função decrescente que assume valores arbitrariamente grandes quando p tende a 0 pela esquerda. Notamos também que quanto menor o valor da constante a mais rapidamente é o crescimento da função à medida que p tende a 0.

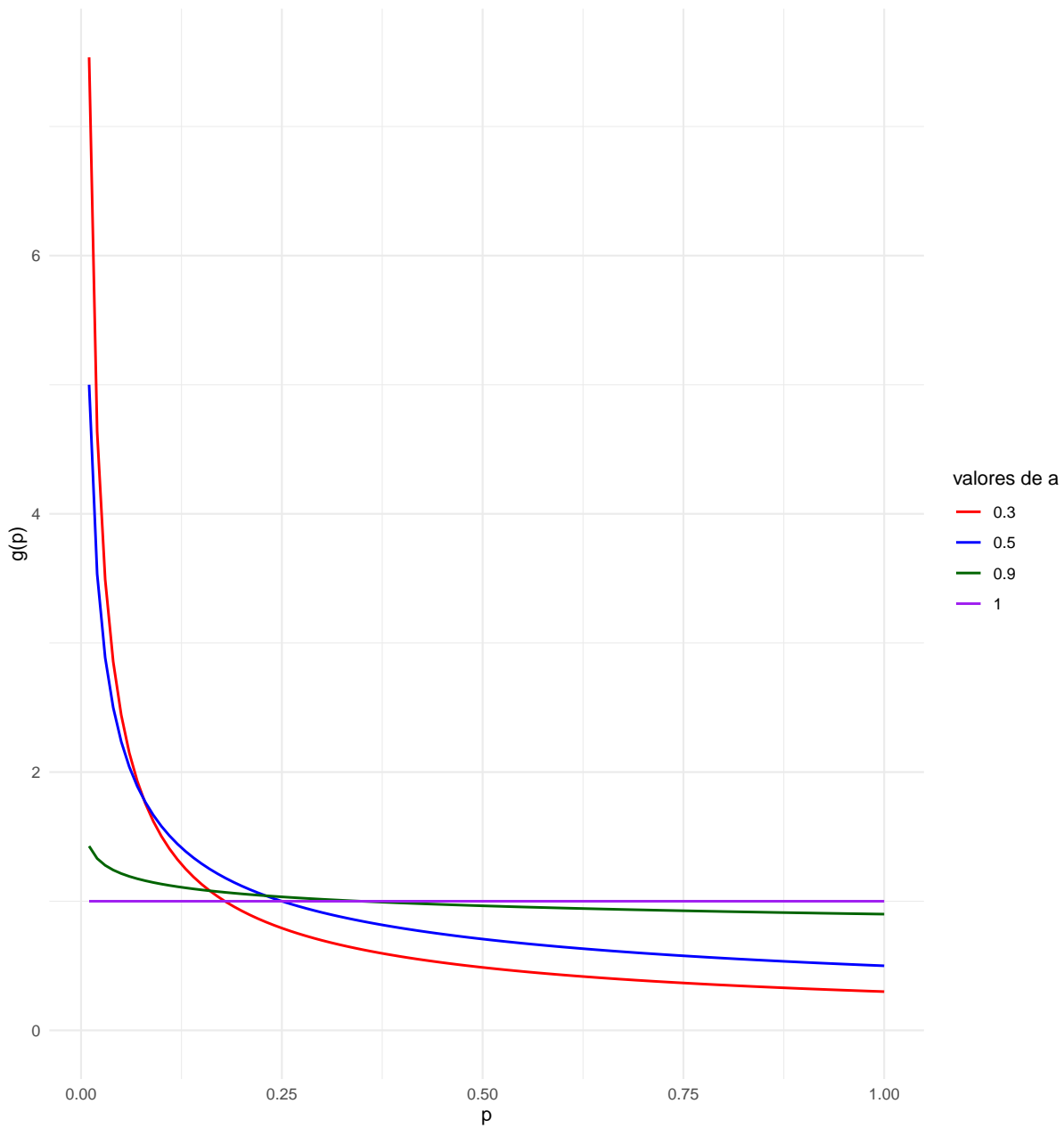


Figura 1 – Gráficos da função g_i dada em (2.15) para os valores de $a = 0,3, 0,5, 0,9$ e 1 .

As expressões para o *exchangeability supermartingale* e para o *exchangeability martingale* definidos em (2.10) e (2.11) são dadas por

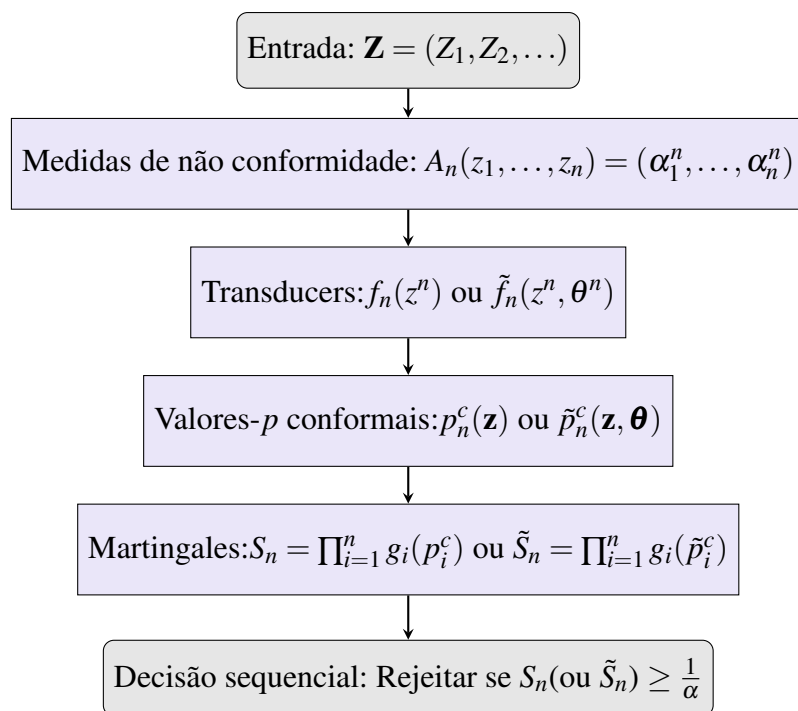
$$S_n = \prod_{i=1}^n a (p_i^c)^{a-1},$$

$$\tilde{S}_n = \prod_{i=1}^n a (\tilde{p}_i^c)^{a-1}.$$

Nos experimentos deste trabalho, utilizamos $a = 0,92$ como sugerido em [Vovk, Nouretdinov e Gammerman \(2003\)](#), que argumentou a respeito de um melhor desempenho na detecção da mudança de distribuição para este valor de a .

2.3.2 Fluxograma resumo

Para facilitar a compreensão do método *Conformal* e a conexão entre seus principais componentes — medidas de não conformidade, *transducers* e martingales —, o fluxograma a seguir foi elaborado como um resumo visual do processo. Ele descreve, de forma concisa, a sequência de etapas envolvidas na construção de testes sequenciais a partir das observações (Z_1, Z_2, \dots) , passando pelo cálculo dos valores- p conformais (determinísticos ou randomizados) até a geração dos martingales responsáveis pela decisão de rejeição da hipótese nula via desigualdade de Ville. Esse esquema serve como guia geral para as seções seguintes e destaca o papel de cada elemento dentro da metodologia.



ANÁLISE EMPÍRICA DA VALIDADE DO MÉTODO *CONFORMAL*

Neste capítulo, são apresentados alguns experimentos realizados para analisar a eficácia das teorias descritas no capítulo anterior e pelos artigos citados do Vovk. A ideia é definir experimentos que validem as pré-suposições do método sob H_0 , assim como realizar uma breve análise dos valores e de como se comportam os Martingales gerados com e sem perturbação. Primeiramente, foram feitas análises sobre os valores p^c originados do método, para testar sua aderência à distribuição uniforme e também sua autocorrelação, assim como foi feita uma análise do controle do erro do tipo I. Além disso, foram geradas observações de diferentes distribuições de probabilidades, realizando perturbações em um determinado momento e analisando o comportamento dos Martingales gerados.

3.1 Análise dos p^c

Neste experimento, são utilizadas duas medidas de não conformidade, sendo denotadas por α^1 e α^2 , que são definidas para cada $z = (z_1, z_2, \dots, z_n)$ como

$$\alpha^k(z) = (\alpha_1^k(z), \alpha_2^k(z), \dots, \alpha_n^k(z)), \quad k = 1, 2,$$

sendo que

$$\alpha_i^1(z) = \min \{d(z_i, z_j) : j \neq i\} \quad (3.1)$$

$$\alpha_i^2(z) = \frac{1}{n-1} \sum_{j=1}^n d(z_i, z_j), \quad (3.2)$$

em que $i = 2, \dots, n$, e $j = 1, \dots, n$ e d é a distância Euclidiana em \mathbb{R}^n .

Resumindo, para cada nova observação na sequência, temos uma atualização da distância mínima (3.1) ou média (3.2) de cada uma dessas observações. Ou seja, está sendo utilizada a mesma ideia de medida de não conformidade apresentada por Vovk (2020), porém, sem levar em conta as classes (pois estamos testando em um contexto diferente de dados univariados).

Nesta seção, é analisado como as pressuposições se comportam, avaliando, sob H_0 , testes de aderência à distribuição uniforme e de autocorrelação dos p^c gerados pelos *transducers*. Para isso, utilizamos duas distribuições, uma contínua, Normal(0,1) e outra discreta, Poisson(2). Desta forma, podemos extrair os p^c através de um *conformal transducer*, utilizando as medidas de não conformidade como *input*.

Durante a simulação, foram geradas **1000 sequências** de observações de vetores de **tamanho 500** a partir de uma distribuição \mathcal{P} . Porém, para análises sob H_0 (sem perturbação), apenas as primeiras 250 observações de cada sequência foram utilizadas. Sendo assim, no primeiro experimento, temos 1000 sequências geradas de uma \mathcal{P}^1 , em que $\mathcal{P}^1 \sim \text{Normal}(0,1)$ e 1000 sequências geradas de uma \mathcal{P}^2 , em que $\mathcal{P}^2 \sim \text{Poisson}(\lambda = 2)$, como mostrado na Tabela 1.

Tabela 1 – Cenários para análise dos valores p^c sob H_0

Cenário	\mathcal{P}
1	Normal(0, 1)
2	Poisson($\lambda = 2$)

Foi realizado o teste Kolmogorov–Smirnov (KS) de aderência quanto à distribuição Uniforme (0,1) para todas as 1000 réplicas. O objetivo é saber a proporção daquelas réplicas em que a hipótese nula de uniformidade foi rejeitada. É esperado que, sendo válido o método, sob H_0 , em média, a fração de réplicas que não são aderentes seja apenas de 5% (no caso dos testes com *transducers* randomizados, vide (2.14)), ou seja, apenas 5% das sequências não seguem uma distribuição uniforme(0,1).

Tabela 2 – Percentual de réplicas nas quais a hipótese nula de uniformidade foi rejeitada no teste KS com $\alpha = 5\%$

Observações	1:250 (H_0)			
	Randomizado		Determinístico	
Medida de não conformidade	α^1	α^2	α^1	α^2
Normal	4.2%	4.8%	11.0%	8.5%
Poisson	4.6%	4.0%	100%	100%

Na Tabela 2, podemos ver a porcentagem de sequências que não seguem uniformidade de acordo com o teste KS. É possível perceber que os valores p^c oriundos de *transducers* determinísticos possuem uma dificuldade maior de seguir uniformidade, principalmente quando os dados são gerados da distribuição discreta de Poisson, em que 100% das sequências foram

rejeitadas quanto à aderência pelo teste KS. Isso é devido ao fato de que esses *transducers* geram valores super-uniformes, como comentado anteriormente. Já para o *transducer* randomizado, temos um comportamento que seria esperado dado que a utilização desse tipo de *transducer* garante teoricamente a uniformidade dos valores p^c gerados, conforme citado no Capítulo 2.

Para corroborar com a análise da tabela, trazemos os gráficos QQ-plots dos valores p^c de cada sequência em relação à Uniforme (0,1), variando a distribuição dos dados, tipo de *transducer* e medida de não conformidade utilizada.

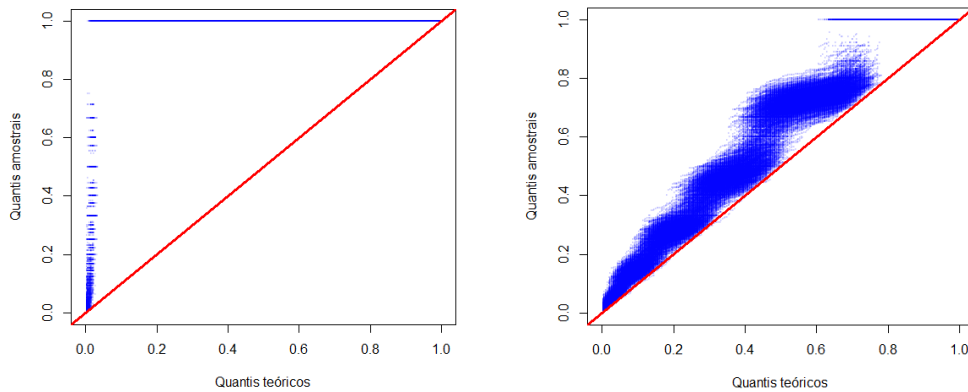


Figura 2 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de **Poisson(2)** utilizando o *transducer* **determinístico**, com α^1 (esquerda) e α^2 (direita).

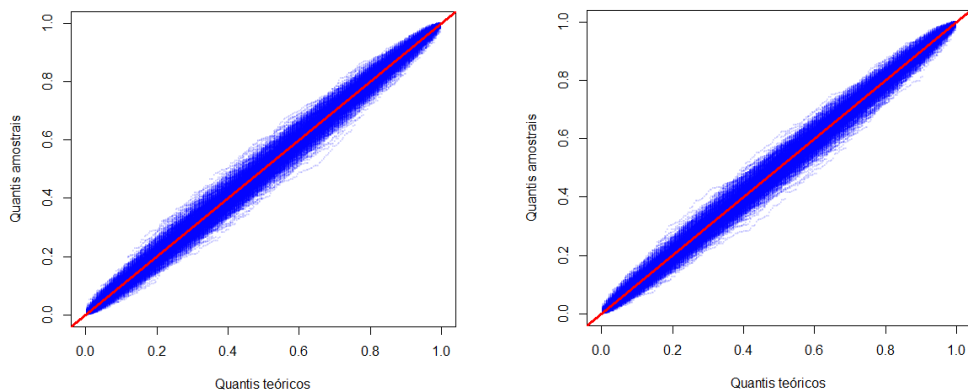


Figura 3 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de **Poisson(2)** utilizando o *transducer* **randomizado**, com α^1 (esquerda) e α^2 (direita).

Por meio das Figuras 3 e 5, e da Tabela 2, podemos perceber que, como esperado, e como explicado na Seção 2.2, a porcentagem de sequências oriundas de *transducers* randomizados apresenta valor conforme o esperado, abaixo de 5%, assim como o *transducer* determinístico não satisfaz propriamente a pressuposição de uniformidade dos p^c , principalmente no cenário da distribuição Poisson.

Além do teste de uniformidade, uma etapa complementar importante na avaliação empírica dos valores- p gerados por *conformal transducers* é a análise de sua independência ao longo

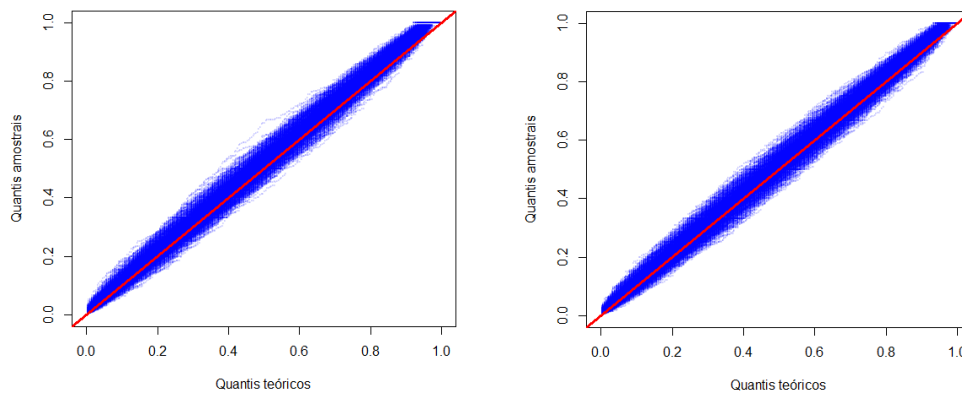


Figura 4 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de **Normal(0,1)** utilizando o *transducer determinístico*, com α^1 (esquerda) e α^2 (direita).

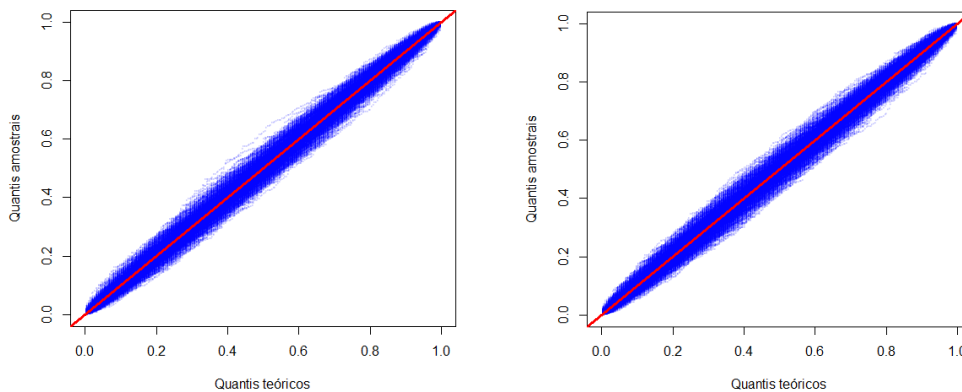


Figura 5 – QQplots para distribuição uniforme dos valores p^c das 1000 réplicas geradas da distribuição de **Normal(0,1)** utilizando o *transducer randomizado*, com α^1 (esquerda) e α^2 (direita).

do tempo. Embora a teoria conformal garanta que, sob permutabilidade, os valores- p sejam marginalmente uniformes (ou superuniformes, no caso determinístico), essa garantia não implica necessariamente independência serial.

Em particular, quando os valores- p são usados para alimentar martingales sequenciais — cuja multiplicação sucessiva é sensível à dependência temporal — a presença de autocorrelação pode afetar o comportamento do processo acumulado. Portanto, a verificação de autocorrelação nas sequências de valores- p contribui para entender se há estruturas temporais residuais, especialmente antes da mudança (isto é, sob a hipótese nula).

Com o intuito de identificar possível dependência nos p^c , foram realizados dois testes de autocorrelação, o teste de Durbin-Watson ([DURBIN; WATSON, 1971](#)) e o teste de Box-Pierce e Ljung-Box, ([BOX; PIERCE, 1970](#); [LJUNG; BOX, 1978](#)).

Tabela 3 – Percentual de réplicas nas quais a hipótese nula de falta de autocorrelação foi rejeitada pelo teste de Durbin-Watson com $\alpha = 5\%$.

Observações <i>Transducer</i>	1:250 (H_0)			
	Randomizado		Determinístico	
Medida de não conformidade	α^1	α^2	α^1	α^2
Normal	5.0%	5.2%	11.5%	9.5%
Poisson	4.9%	4.7%	31.0%	6.9%

Tabela 4 – Percentual de réplicas nas quais a hipótese nula de falta de autocorrelação foi rejeitada pelo teste de Box-Pierce-Ljung-Box com $\alpha = 5\%$.

Observações <i>Transducer</i>	1:250 (H_0)			
	Randomizado		Determinístico	
Medida de não conformidade	α^1	α^2	α^1	α^2
Normal	4.7%	4.9%	5.4%	5.6%
Poisson	5.4%	4.7%	26.7%	4.5%

Levando em consideração os resultados dos testes apresentados nas Tabelas 3 e 4, percebe-se que tanto para a distribuição Normal quanto para a Poisson, os p^c gerados pelo *transducer* randomizado apresentam resultados com menor auto-correlação, conforme o esperado. Em relação ao *transducer* determinístico, a ausência de autocorrelação foi rejeitada acima do esperado (5%) em sua totalidade (tipo de distribuição e medida de não conformidade) utilizando o teste de Durbin-Watson, porém apresentou um comportamento um pouco melhor pelo teste de Box-Pierce-Ljung-Box.

3.2 Análise dos Martingales.

Nesta seção, são apresentados os experimentos realizados com os valores dos martingales obtidos para quatro distribuições de probabilidade. Assim como no experimento anterior, foram geradas 1000 sequências de tamanho 500, nas quais as primeiras 250 observações foram geradas a partir de uma distribuição \mathcal{P}^1 e as outras 250 observações na sequência, de uma distribuição \mathcal{P}^2 . Foram escolhidas algumas distribuições específicas, sendo essas a Normal, Gamma, Exponencial e Poisson, obtendo os cenários apresentados na Tabela 5. O objetivo foi realizar perturbações bem significativas para analisar o comportamento dos martingales.

Tabela 5 – Cenários para análise dos Martingales

Cenário	\mathcal{P}_1	\mathcal{P}_2
1	Normal(0, 1)	Normal(10, 25)
2	Gamma($k = 2, \theta = 0.1$)	Gamma($k = 5, \theta = 1$)
3	Exp(20)	Exp(0.01)
4	Pois($\lambda = 2$)	Pois($\lambda = 20$)

3.2.1 Controle do Erro do Tipo I

Primeiramente, é realizada a análise do controle do erro do tipo I para os valores de cada martingale, utilizando os primeiros 250 valores das sequências (pré-perturbação). Ou seja, alternando os valores do nível de significância (e conseqüentemente o valor limite de detecção C , baseado em (2.14)), o objetivo é avaliar o quão bem cada combinação de martingale, medida de não conformidade e distribuição se comporta. Sendo assim, é medida, sob H_0 , a porcentagem de sequências com alarme falso para diversos níveis de significância α , $0 \leq \alpha \leq 1$.

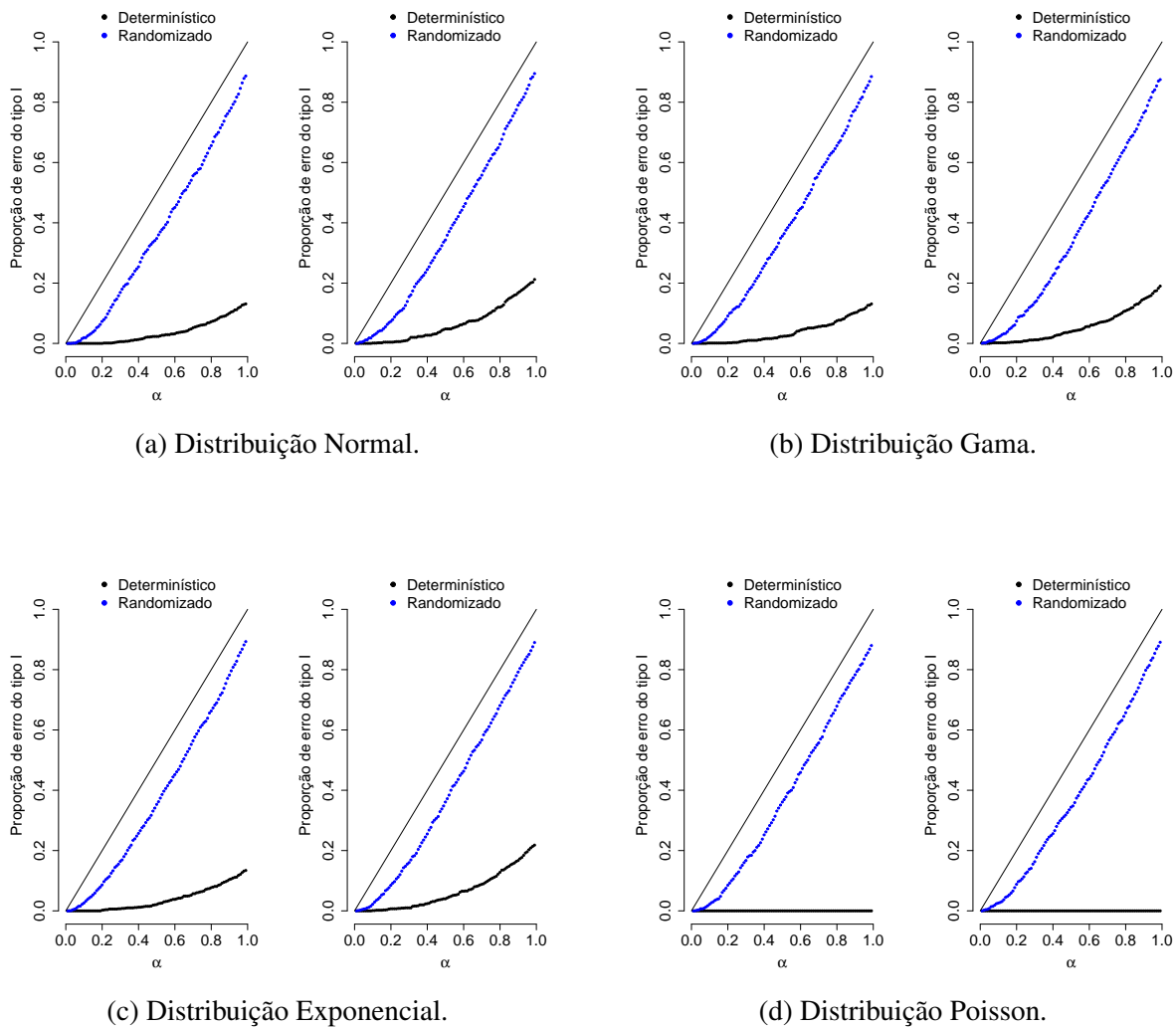


Figura 6 – Proporção de falsos alarmes (erro tipo I) para as diferentes distribuições do cenário descrito na tabela 5, em que à esquerda de cada par de gráficos, temos a medida de não conformidade α^1 e à direita, α^2

Os gráficos apresentados nas Figuras 6a, 6b, 6c e 6d mostram o controle do erro tipo I de cada método, comparando o desempenho dos *transducers* randomizados e determinísticos. Este resultado implica que os testes baseados em *transducers* randomizados têm necessariamente maior poder que aqueles baseados em *transducers* determinísticos.

Entretanto, no caso da **distribuição de Poisson**, o *transducer* determinístico apresentou um número extremamente baixo de falsos positivos, em alguns casos nenhum falso positivo. Embora isso possa parecer um bom resultado à primeira vista, na prática, indica um comportamento **excessivamente conservador**, diminuindo o poder de detecção de mudanças quando elas ocorrem.

Esse comportamento sugere que *transducers* randomizados podem ser mais eficazes em cenários em que o controle do erro tipo I precisa ser ajustado de maneira mais próxima ao valor esperado α , garantindo um equilíbrio adequado entre falsos positivos e falsos negativos. Isso é especialmente relevante para aplicações como monitoramento de sistemas, segurança financeira e detecção de fraudes, onde um excesso de falsos negativos pode ser tão problemático quanto um número elevado de falsos positivos.

3.2.2 Análise dos valores dos martingales.

Nas Figuras 7, 8, 9 e 10 temos os valores de cada martingale apresentado na escala log (para melhor visualização) e o respectivo limiar $\log(C) \approx 3$ de detecção baseado na inequação de Ville (2.14) (com $C = 20$, de acordo com $\alpha = 5\%$). Esses gráficos nos permitem analisar intuitivamente os valores destes martingales. Nestes gráficos em específico, os martingales, tanto \tilde{S}_n como S_n , estão definidos como M_n no eixo y.

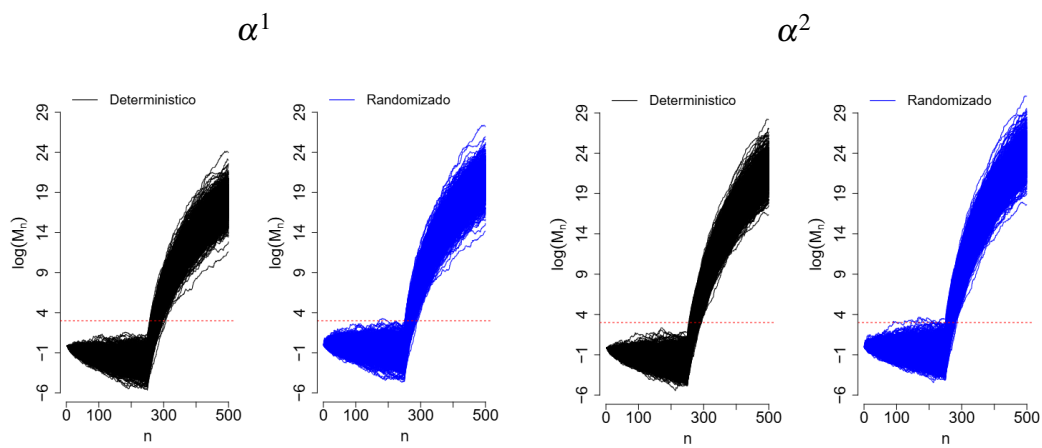


Figura 7 – Martingales do Experimento 1 (Normal), à esquerda, gerados por α^1 e à direita, por α^2 .

As Figuras 7,8,9 e 10 mostram que assim que a perturbação ocorre ($n = 250$), os martingales baseados no *transducer* randomizado disparam para cima em todas as distribuições (Normal, Gamma, Exponencial e Poisson com α^2). Isso se deve ao fato de que, antes da mudança, tais valores- p são genuinamente uniformes; depois da quebra da permutabilidade, passam a produzir muitos valores pequenos, ocasionando acúmulos cada vez maiores nos martingales, dada sua construção.

O comportamento do *transducer* determinístico é diferente: para Poisson com α^1 (Figura 10, primeiro painel) embora o processo cresça brevemente após a perturbação, logo depois acaba

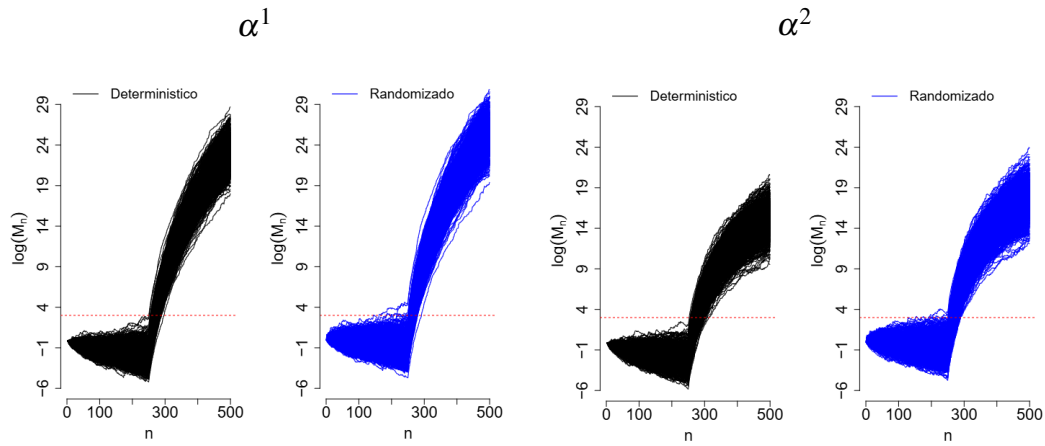


Figura 8 – Martingales do Experimento 2 (Gamma), à esquerda, gerados por α^1 e à direita, por α^2 .

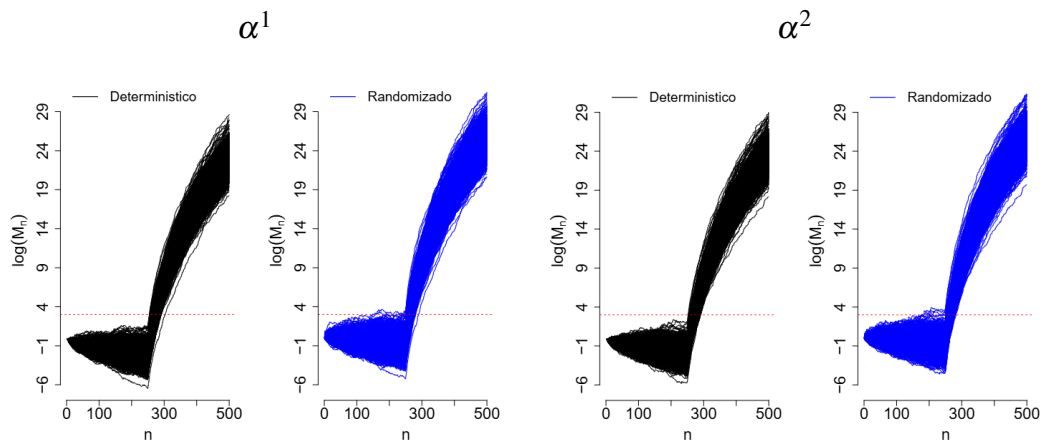


Figura 9 – Martingales do Experimento 3 (Exponencial), à esquerda, gerados por α^1 e à direita, por α^2 .

caindo continuamente. A explicação é que, neste caso, os valores- p permanecem conservativos e o processo gerado é apenas um *supermartingale*, de acordo com o Teorema 3, o que faz seu valor tender a zero quando não surgem valores- p suficientemente pequenos. Assim, embora evite alarmes falsos (erro tipo I), esse *transducer* torna-se demasiado conservador, perdendo

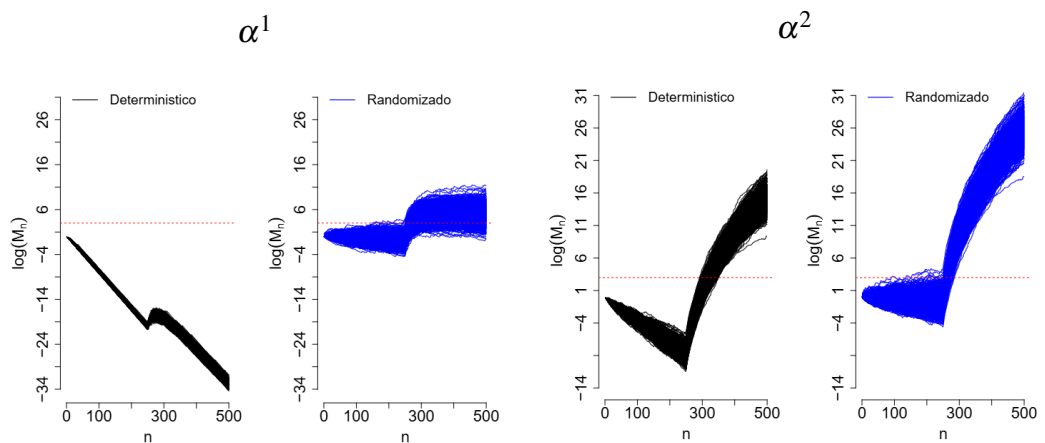


Figura 10 – Martingales do Experimento 4 (Poisson), à esquerda, gerados por α^1 e à direita, por α^2 .

potência para detectar a mudança (erro tipo II elevado). Em contraste, o uso de α^2 atenua a super-uniformidade dos p -valores, restabelecendo o crescimento do martingale e aumentando a sensibilidade ao *shift*. Esses resultados reforçam que a eficácia do teste sequencial depende da combinação medida de não-conformidade + tipo de *transducer*, devendo-se balancear falsos positivos e falsos negativos segundo a aplicação.

COMPARAÇÃO EMPÍRICA ENTRE MÉTODOS CLÁSSICOS E MÉTODOS BASEADOS EM *CONFORMAL* MARTINGALES

Este capítulo apresenta uma análise comparativa entre métodos clássicos de detecção de pontos de mudança, como CUSUM (PAGE, 1954), Shiryaev-Roberts (SHIRYAEV, 1961) e *Posterior Probability* (SHIRYAEV, 1963) e métodos baseados nos *conformal* martingales apresentados anteriormente, com diferentes medidas de não conformidade α^i sendo utilizadas. Em (VOLKHONSKIY *et al.*, 2017) foram realizadas comparações neste sentido, porém utilizando apenas duas medidas de não conformidade. Neste trabalho, em um primeiro momento, são realizadas comparações em cenários cujas distribuições pré e pós-mudança (isto é, $\mathcal{P}^{(0)}$ e $\mathcal{P}^{(1)}$) são conhecidas pelos métodos clássicos, e posteriormente, é realizado um teste em que é feita a especificação errada dessas distribuições, pois como dito na Introdução 1.1, os métodos clássicos se baseiam na especificação exata destas distribuições, o que acaba sendo uma limitação. Em ambos os casos, testamos tanto distribuições univariadas, como na Seção 4.2, como para casos multivariados, na Seção 4.3, onde é apresentada uma gama de novas medidas de não conformidade sendo utilizadas.

Em seguida, na Seção 4.4, apresenta-se um experimento adicional no qual se demonstra o impacto de fornecer informações incorretas sobre a distribuição pós-mudança aos métodos clássicos, evidenciando uma limitação importante dessas abordagens.

Antes de apresentarmos os métodos clássicos de detecção de mudança, introduzimos a notação e o contexto adotados neste capítulo.

Seja $n \geq 1$ o número total de observações, f_0 e f_1 as funções densidade de probabilidade (fdp's) antes e depois da mudança, isto é, z_1, \dots, z_n são variáveis independentes tais que:

$$z_1, \dots, z_{i-1} \sim f_0, \quad z_i, \dots, z_n \sim f_1,$$

para algum $i = 1, \dots, n$. Definimos as funções de verossimilhança:

$$L_n^i = \prod_{j=1}^{i-1} f_0(z_j) \prod_{j=i}^n f_1(z_j), \quad L_n = \prod_{j=1}^n f_0(z_j).$$

A verossimilhança L_n^i representa a probabilidade ou densidade de (z_1, \dots, z_n) supondo que $i \in \{1, \dots, n\}$ é o ponto de mudança, enquanto que L_n é a fdp de (z_1, \dots, z_n) assumindo que não houve mudança.

4.1 Métodos clássicos

Os métodos clássicos de detecção de pontos de mudança, como **CUSUM**, **Shiryaev-Roberts (S-R)** e o método **Posterior Probability (PP)** — este último com uma abordagem bayesiana — assumem que a densidade de probabilidade das observações muda de uma distribuição conhecida \mathcal{P}^0 para outra \mathcal{P}^1 em um ponto de mudança $\theta \in \{1, \dots, n\}$ desconhecido.

Essas abordagens dependem de informações paramétricas completas sobre as distribuições antes e depois da mudança. Além disso, cada método define uma estatística específica — como γ_n , ψ_n , ou ϕ_n — cuja função é acumular evidência contra a hipótese de que os dados seguem a distribuição original. Valores altos dessas estatísticas são interpretados como indícios de que uma mudança pode ter ocorrido, servindo assim como critério de parada da detecção.

4.1.1 CUSUM (Cumulative Sum):

O método CUSUM, introduzido por (PAGE, 1954), é um dos mais clássicos e amplamente utilizados para detecção sequencial de mudanças. Ele se baseia na razão de verossimilhança acumulada das observações, e sua regra de detecção é definida como:

$$\tau_{\text{CUSUM}} = \inf \{n : \gamma_n \geq C\},$$

cuja a estatística γ_n é dada por:

$$\gamma_n = \max_{\theta \in [1, n]} \log \left(\frac{L_n^\theta}{L_n} \right) = \max_{1 \leq i \leq n} \sum_{j=i}^n \log (R(z_j)),$$

com $R(z_j) = \frac{f_1(z_j)}{f_0(z_j)}$ sendo a razão de verossimilhança entre os modelos com e sem mudança.

Essa estatística equivale à razão de verossimilhança entre o modelo com mudança em i e o modelo sem mudança, escrita em escala logarítmica. A intuição é que, a partir de um certo

ponto i , os dados passam a ser mais prováveis sob f_1 do que sob f_0 , e o método busca identificar esse ponto de forma eficiente.

Para facilitar a implementação computacional, pode-se utilizar a forma recursiva da estatística CUSUM (dedução apresentada no Apêndice A.2):

$$\begin{cases} \gamma_1 = \log(R(z_1)), \\ \gamma_{n+1} = \log(R(z_{n+1})) + \max(\gamma_n, 0), \quad n \geq 1. \end{cases}$$

Este método é considerado ótimo sob o critério minimax, conforme demonstrado por (MOUSTAKIDES, 1986), que provou que a regra de parada CUSUM minimiza o pior atraso de detecção esperado, dado um nível fixo de falsos alarmes. No contexto de detecção sequencial, esse critério busca minimizar o maior atraso de detecção esperado, independentemente do instante θ em que a mudança ocorre. Isso significa que, mesmo nos cenários mais adversos, o método garante um limite superior para o tempo de detecção.

4.1.2 Shiryaev-Roberts (SR):

Proposto originalmente por (SHIRYAEV, 1961) e posteriormente analisado por (ROBERTS, 1966), o método Shiryaev–Roberts (SR) busca minimizar o atraso de detecção esperado, sendo assintoticamente ótimo sob certos critérios no contexto minimax (POLLAK, 1985).

Sua regra de detecção é definida por:

$$\tau_{SR} = \inf \{n : \psi_n \geq C\},$$

cuja a estatística ψ_n é definida como:

$$\psi_n = \log \left(\sum_{\theta=1}^n \frac{L_n^\theta}{L_n} \right) = \log \left(\sum_{i=1}^n \prod_{j=i}^n R(z_j) \right),$$

com $R(z_j) = \frac{f_1(z_j)}{f_0(z_j)}$ sendo a razão de verossimilhança entre os modelos com e sem mudança.

Esta expressão pode ser reescrita de forma recursiva, o que facilita sua implementação computacional. A forma recursiva é dada por (dedução apresentada em A.3):

$$\begin{cases} \psi_1 = \log(R(z_1)), \\ \psi_{n+1} = \log(R(z_{n+1})) + \log(e^{\psi_n} + 1), \quad n \geq 1. \end{cases}$$

Essa formulação permite que a estatística seja atualizada iterativamente, sem necessidade de armazenar todo o histórico de observações.

4.1.3 Probabilidade Posterior (PP):

No cenário bayesiano, (SHIRYAEV, 1963) propôs o uso da probabilidade a posteriori para detectar mudanças. A ideia é calcular, a cada instante n , a probabilidade de que a mudança tenha ocorrido em algum ponto $\theta \in \{1, \dots, n\}$, dado os dados observados até o instante n .

A regra de detecção é definida como:

$$\tau_{PP} = \inf \{n : \phi_n \geq C\},$$

em que

$$\phi_n = \log \left(\frac{\sum_{\theta=1}^n L_n^\theta \pi(\theta)}{L_n (1-p)^n} \right) = \log \left(\frac{\sum_{\theta=1}^n \left(\prod_{j=\theta}^n R(z_j) \right) \pi(\theta)}{(1-p)^n} \right),$$

com $\pi(\theta) = p(1-p)^{\theta-1}$ sendo a distribuição a priori geométrica para o ponto de mudança, L_n^θ a verossimilhança assumindo mudança no tempo θ , e L_n a verossimilhança sob a hipótese nula de nenhuma mudança. Neste trabalho, fixamos $p = 0.01$, valor já utilizado por outros artigos anteriormente.

Para propósitos computacionais, pode-se utilizar a forma recursiva da estatística ϕ_n , dada por (dedução apresentada no Apêndice A.1):

$$\begin{cases} \phi_1 = \log(p) - \log(1-p) + \log(R(z_1)), \\ \phi_{n+1} = \log(R(z_{n+1})) - \log(1-p) + \log(p + e^{\phi_n}), \quad n \geq 1, \end{cases}$$

em que $R(z_j) = \frac{f_1(z_j)}{f_0(z_j)}$ é a razão de verossimilhança ponto a ponto.

Essa formulação permite atualizar a estatística de forma eficiente em tempo real, sendo usada nos experimentos computacionais deste trabalho.

4.2 Experimentos Univariados

Para avaliar o desempenho dos métodos de detecção de mudança, foram realizados experimentos controlados com dados simulados. Cada cenário foi **replicado 1000 vezes**, gerando sequências univariadas de tamanho $n = 500$, com uma mudança estrutural introduzida na observação $\theta = 250$ (ou seja, na metade da sequência). Os experimentos consideram três distribuições distintas para as observações: Normal, Poisson e Exponencial. Em cada caso, os parâmetros da distribuição \mathcal{P}^1 foram ajustados para representar diferentes intensidades de mudança em relação à distribuição original \mathcal{P}^0 , conforme detalhado na Tabela 6. As distribuições selecionadas refletem diferentes características estatísticas relevantes para a avaliação da robustez dos métodos. A distribuição Normal representa um cenário contínuo e simétrico, amplamente utilizado como referência em testes paramétricos. A Poisson, por outro lado, é discreta e assimétrica, adequada

para modelar contagens. Já a Exponencial representa o caso contínuo e assimétrico, com cauda longa, frequentemente usado para modelar tempos entre eventos.

Os parâmetros escolhidos para \mathcal{P}^0 e \mathcal{P}^1 foram definidos de forma a garantir uma variação controlada e progressiva na magnitude da mudança. Para a Normal, apenas a média foi alterada, mantendo a variância fixa, o que permite isolar o impacto da translação. Na Poisson e na Exponencial, o valor de λ foi manipulado para provocar mudanças perceptíveis na dispersão e na cauda da distribuição. Essa escolha permite observar como os métodos respondem a diferentes tipos e intensidades de perturbação, abrangendo cenários com estruturas probabilísticas contínuas e discretas, leves e abruptas, simétricas e assimétricas.

Tabela 6 – Distribuições utilizadas nos experimentos

Distribuição	Parâmetros iniciais (\mathcal{P}^0)	Parâmetros após mudança (\mathcal{P}^1)
Normal	$\mu_0 = 0, \sigma_0 = 1$	$\mu_1 \in \{1, 1.5, 2\}, \sigma_1 = \sigma_0$
Poisson	$\lambda_0 = 1$	$\lambda_1 \in \{2, 4, 10\}$
Exponencial	$\lambda_0 = 20$	$\lambda_1 \in \{5, 2, 1\}$

Os métodos avaliados incluem os modelos clássicos paramétricos (CUSUM, Shiryaev–Roberts e *Posterior Probability*) e os métodos baseados em *conformal* martingales, os quais não requerem conhecimento prévio das distribuições antes e depois da mudança. Para os métodos *conformal*, foram utilizadas duas medidas de conformidade distintas, α^1 , definida na Eq. (3.1) e α^2 , definida na Eq. (3.2).

O desempenho de cada método foi avaliado com base em uma curva construída a partir da variação do parâmetro de parada C , que regula a sensibilidade da detecção. Para cada valor de C , calcula-se:

- A **taxa de falso positivo**, $P_0(\tau \leq \theta)$, que representa a proporção de réplicas sob a hipótese nula (sem mudança), nas quais o alarme foi disparado **antes** do ponto θ ;
- O **tempo médio de detecção**, $\mathbb{E}_1[\tau - \theta \mid \tau \geq \theta]$, que corresponde ao atraso médio observado nas réplicas, sob a hipótese alternativa (com mudança), considerando apenas os casos em que a mudança foi corretamente detectada.

Esses dois valores formam um par (x, y) para cada C , onde $x = P_0(\tau \leq \theta)$ e $y = \log_{10}(1 + \mathbb{E}_1[\tau - \theta \mid \tau \geq \theta])$, em que y foi definido dessa forma para lidar com escalas muito diferentes de tempos de detecção, evitar indefinições numéricas, e permitir visualização e comparação mais claras entre métodos. Ao variar C dentro de uma faixa predefinida (escolhida a partir dos valores observados no conjunto de dados simulados, de modo a cobrir adequadamente a amplitude dos valores gerados), obtém-se uma sequência de pontos que define uma curva de desempenho para o método.

Esse gráfico permite observar o seguinte *trade-off*, quando C é alto, há menos falsos positivos, mas a detecção tende a ocorrer com mais atraso, quando C é baixo, a detecção é rápida, mas ao custo de maior taxa de falsos positivos.

Para comparar os métodos de forma agregada, foi utilizada uma métrica de área sob a curva, que resume o desempenho da abordagem ao longo de toda a faixa de sensibilidade. A área sob a curva (AUC) resume a curva de desempenho em um único número, permitindo a comparação quantitativa entre métodos. Quanto menor a área, melhor o desempenho geral — já que valores baixos de y (tempo de detecção) combinados com baixos x (taxa de falso positivo) indicam métodos eficazes e confiáveis. Definimos:

$$\text{AUC} \approx \sum_{i=1}^{n-1} (x_{i+1} - x_i) \cdot \frac{y_i + y_{i+1}}{2},$$

em que (x_i, y_i) são os pares ordenados de taxa de falso positivo e tempo de detecção.

Para garantir a comparabilidade entre os métodos, foi realizada uma interpolação linear dos pares $(P_0(\tau \leq \theta), \log_{10}(1 + \mathbb{E}[(\tau - \theta) | \tau \geq \theta]))$ de cada método sobre uma grade comum de valores de $P_0(\tau \leq \theta)$, definida no intervalo $[0, 0.8]$ com espaçamento de 0.001.

Essa padronização é necessária porque, originalmente, cada método gera suas curvas de desempenho em posições distintas no eixo x (taxa de falso positivo), dependendo da sua dinâmica interna de detecção para diferentes limiares C . Alguns métodos podem ter muitos pontos concentrados em certas regiões do eixo x , enquanto outros podem cobrir faixas diferentes ou ter menos pontos computados.

Ao aplicar a interpolação linear, todos os métodos passam a ser avaliados exatamente nos mesmos pontos x , assegurando que a integral (área sob a curva, AUC) seja calculada sobre um domínio compartilhado. Isso evita que um método seja favorecido ou prejudicado artificialmente devido à maior ou menor densidade de pontos em regiões específicas do gráfico.

Em resumo, a interpolação linear uniformiza a base de comparação, permitindo que a área sob a curva reflita exclusivamente as diferenças de desempenho entre os métodos, e não diferenças nos pontos de amostragem.

Por fim, para permitir a comparação das AUCs entre diferentes métodos e cenários, as áreas calculadas foram normalizadas por uma área máxima teórica A_{\max} . Essa área representa o limite superior que a curva poderia atingir se apresentasse, em toda a extensão de x , o maior valor observado de desempenho (tempo médio de detecção mínimo, ou seja, y_{\min}).

Formalmente, como a curva está definida sobre o eixo $x = P_0(\tau \leq \theta)$, variando de um valor mínimo x_{\min} até um máximo x_{\max} (definidos por uma grade de valores entre 0 e 1), e o melhor desempenho observado ocorre quando $y(x)$ é constante e igual ao menor tempo de detecção possível, então a área máxima teórica é dada por:

$$A_{\max} = \int_{x_{\min}}^{x_{\max}} \max(y) dx = (x_{\max} - x_{\min}) \cdot \max(y),$$

em que $\max(y)$ corresponde ao maior valor observado no eixo y , isto é, o pior tempo de detecção médio entre todos os métodos considerados.

Essa normalização posiciona a AUC em uma escala relativa entre 0 e 1, permitindo comparações justas entre experimentos com diferentes densidades de pontos, limites do eixo x , ou escalas de tempo de detecção. Além disso, evita que curvas mais extensas artificialmente obtenham áreas maiores apenas por cobrirem um intervalo maior de $P_0(\tau \leq \theta)$.

Nota sobre a notação das figuras: Nos gráficos apresentados ao longo deste e dos próximos capítulos, os rótulos “ $\alpha = 1$ ” e “ $\alpha = 2$ ” se referem, respectivamente, às medidas de não conformidade α^1 e α^2 conforme definidas em cada capítulo. Assim como os rótulos M se referem aos martingales, sendo $MDet = S_n$ e $MRnd = \tilde{S}_n$. Essas simplificações na legenda foram adotadas por praticidade no código, mas não altera a interpretação dos resultados discutidos.

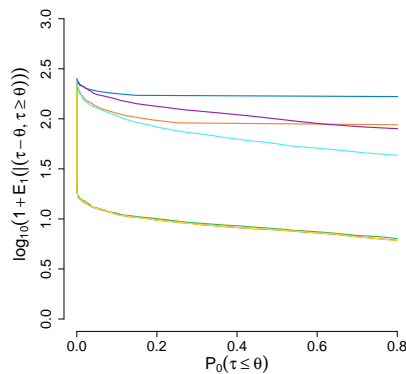
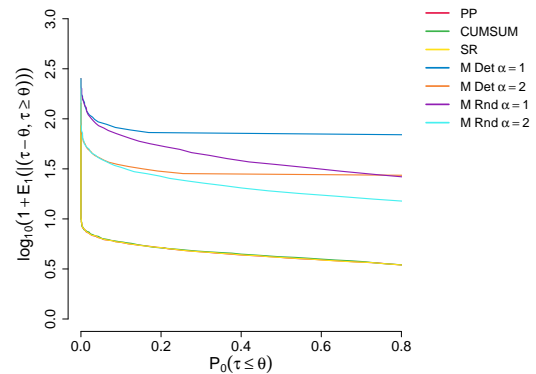
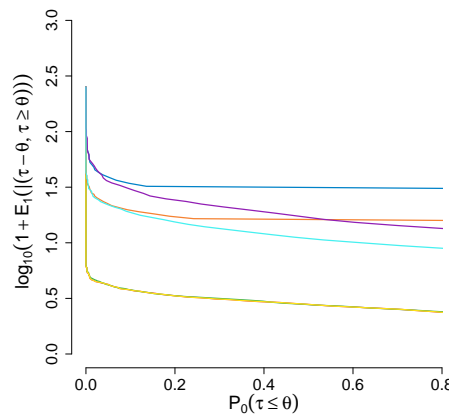
(a) $\mu_1 = 1$ (b) $\mu_1 = 1.5$ (c) $\mu_1 = 2$

Figura 11 – Comparação dos métodos para distribuição Normal, com diferentes valores para μ_1 .

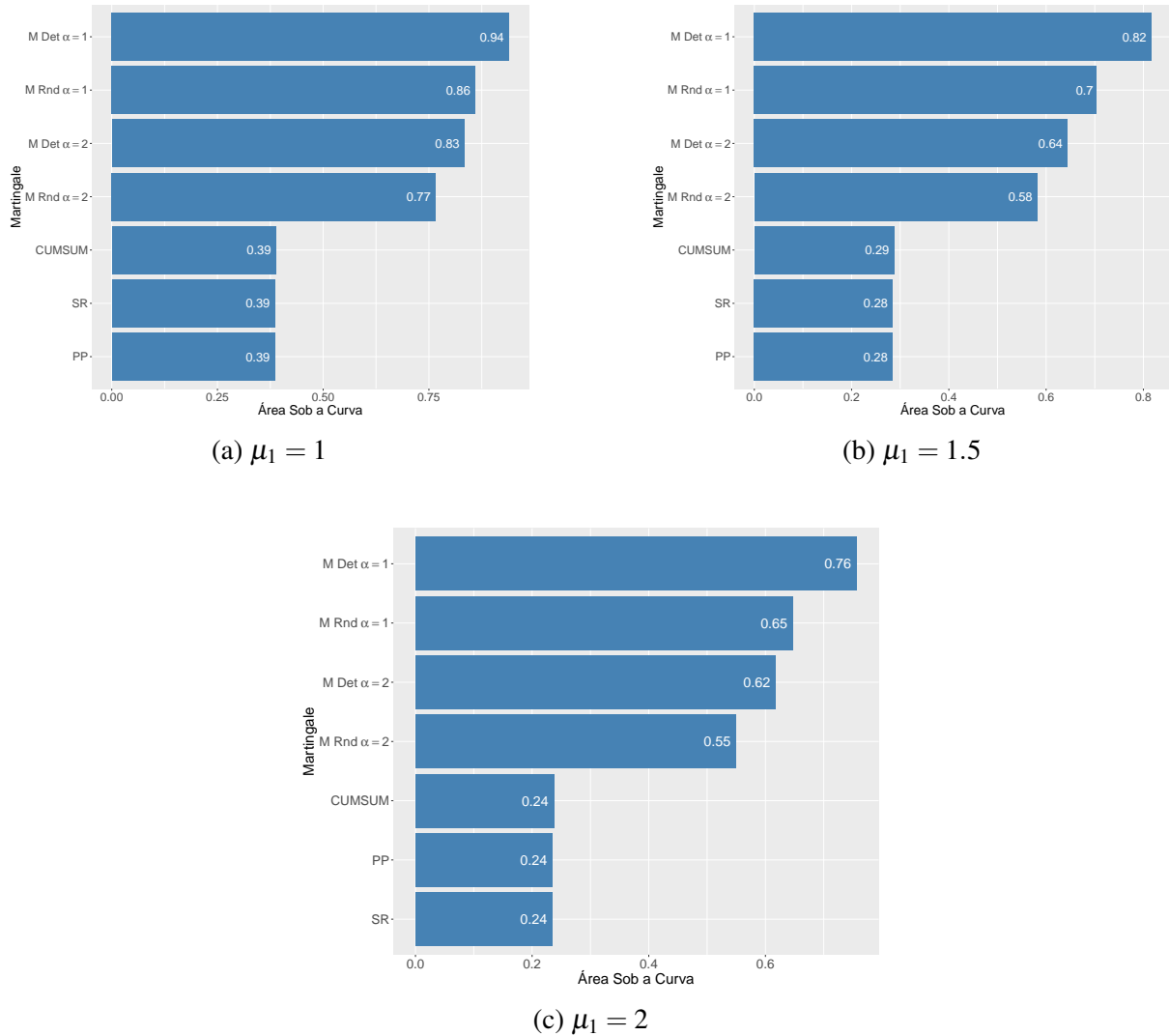


Figura 12 – AUC para distribuição Normal, com diferentes valores para μ_1 .

Observa-se que, em todos os experimentos, os métodos clássicos apresentam curvas sistematicamente inferiores – isto é, com tempos médios de detecção menores para os mesmos níveis de falso alarme – conforme ilustrado nas Figuras 11, 13 e 15, com as curvas de comparação, e nas Figuras 12, 14 e 16, com as respectivas áreas abaixo das curvas. Esse desempenho superior deve-se à especificação correta das distribuições pré e pós-mudança, condição que, apesar de proporcionar detecções mais rápidas, é frequentemente difícil de obter na prática. Outra observação é que todos os métodos são sensíveis à grandeza da perturbação, pois, conforme os parâmetros de \mathcal{P}^1 vão ficando mais distantes dos parâmetros de \mathcal{P}^0 , os métodos vão detectando mais rapidamente as mudanças.

Em geral, a versão com *transducer* randomizado tende a alcançar detecções ligeiramente mais rápidas do que a determinística, conforme pode ser visualizado, por exemplo, na Figura 11 (Normal) e na Figura 15 (Exponencial). Entretanto, para a distribuição Poisson (Figuras 13 e 14), as curvas dos martingales determinísticos permanecem praticamente planas – devido à ausência de alarmes falsos – enquanto os martingales originados de *transducers* randomizados apresentam

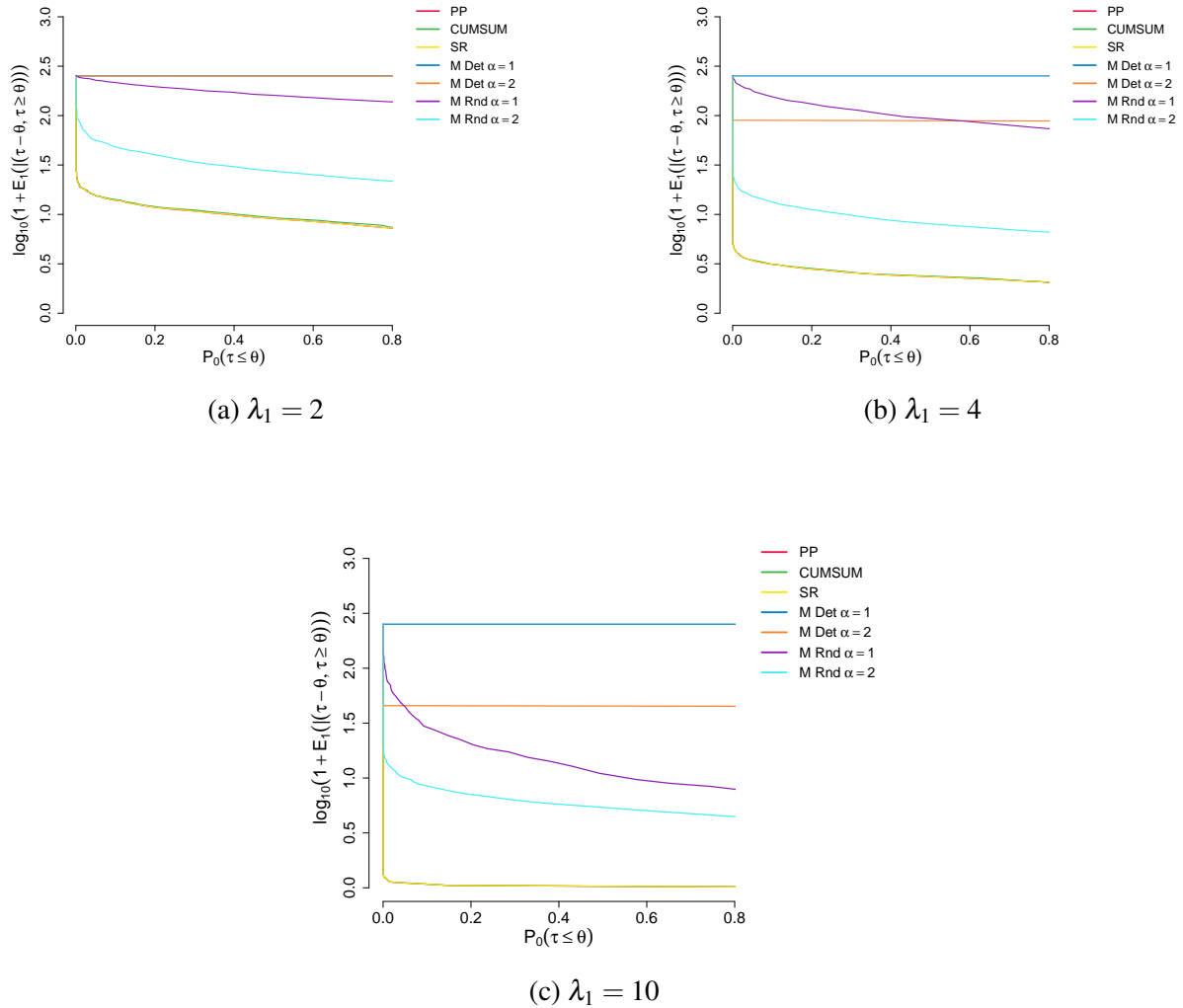


Figura 13 – Comparação dos métodos para distribuição Poisson, com diferentes valores para λ_1 .

desempenho intermediário. Essa característica – ausência de alarmes falsos combinada com detecções muito tardias – ocasionou a exclusão dos respectivos AUCs na Figura 14, reforçando a necessidade de interpretar esses resultados com cautela.

A comparação entre os valores α^1 e α^2 (como destacado nas subdivisões internas de cada figura) evidencia que, com α^2 , os martingales tendem a ter um desempenho melhor, detectando a mudança um pouco mais cedo, embora isso possa ocorrer às custas de um controle menos rígido dos alarmes falsos. Com α^1 , os métodos se mostram mais conservadores, resultando em curvas um pouco superiores, ou seja, maior tempo médio de detecção.

De maneira geral, os resultados apontam que os métodos clássicos demonstram desempenho superior em termos de tempo médio de detecção, conforme evidenciado pelas AUCs nas Figuras 12, 14 e 16.

Lembramos que, embora os *conformal* martingales apresentem atrasos maiores, eles oferecem a vantagem de não requerer a especificação completa das distribuições, o que pode ser

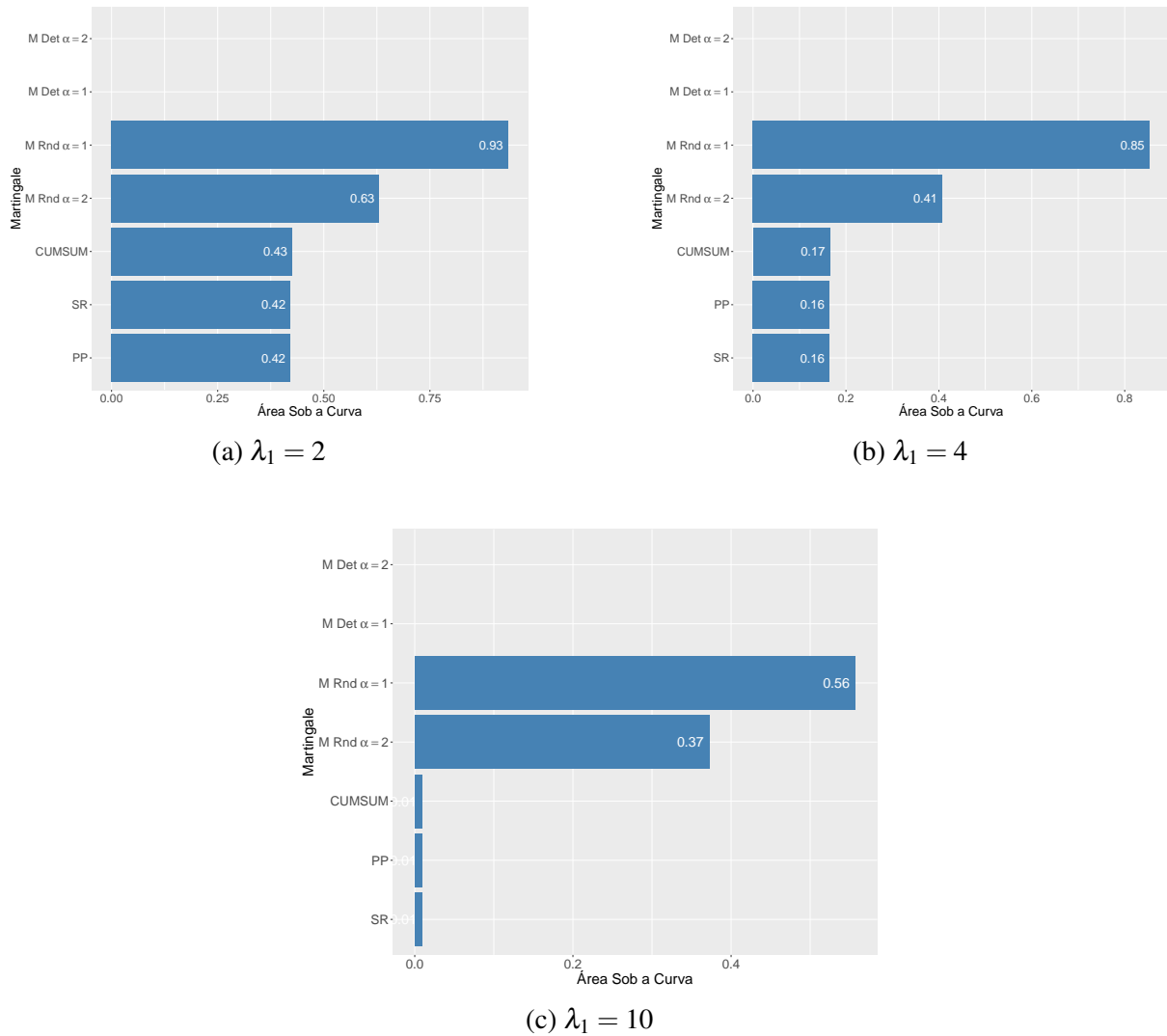


Figura 14 – AUC para distribuição Poisson, com diferentes valores para λ_1 .

determinante em cenários práticos. Esse fato será posteriormente abordado nas próximas seções.

Entre os métodos baseados em *conformal* martingales, a abordagem randomizada tende a equilibrar melhor a rapidez de detecção e o controle dos falsos alarmes, enquanto a versão determinística se mostra menos eficaz, especialmente na distribuição Poisson.

4.3 Experimentos Multivariados

Expandimos as análises anteriores para o caso multivariado, inicialmente na Subseção 4.3.1 são realizados experimentos para versões multivariadas da distribuição Normal com $d = 2$, utilizando as medidas de não conformidade α^1 e α^2 (Eq. 3.1 e Eq. 3.2) adaptadas para o caso multivariado. Em seguida na Subseção 4.3.2, é abordado ainda o caso multivariado da Normal, porém com mais dimensões ($d = 10$) e são introduzidas novas medidas de não conformidade, baseadas no conceito de profundidade estatística, utilizando várias medidas de profundidade

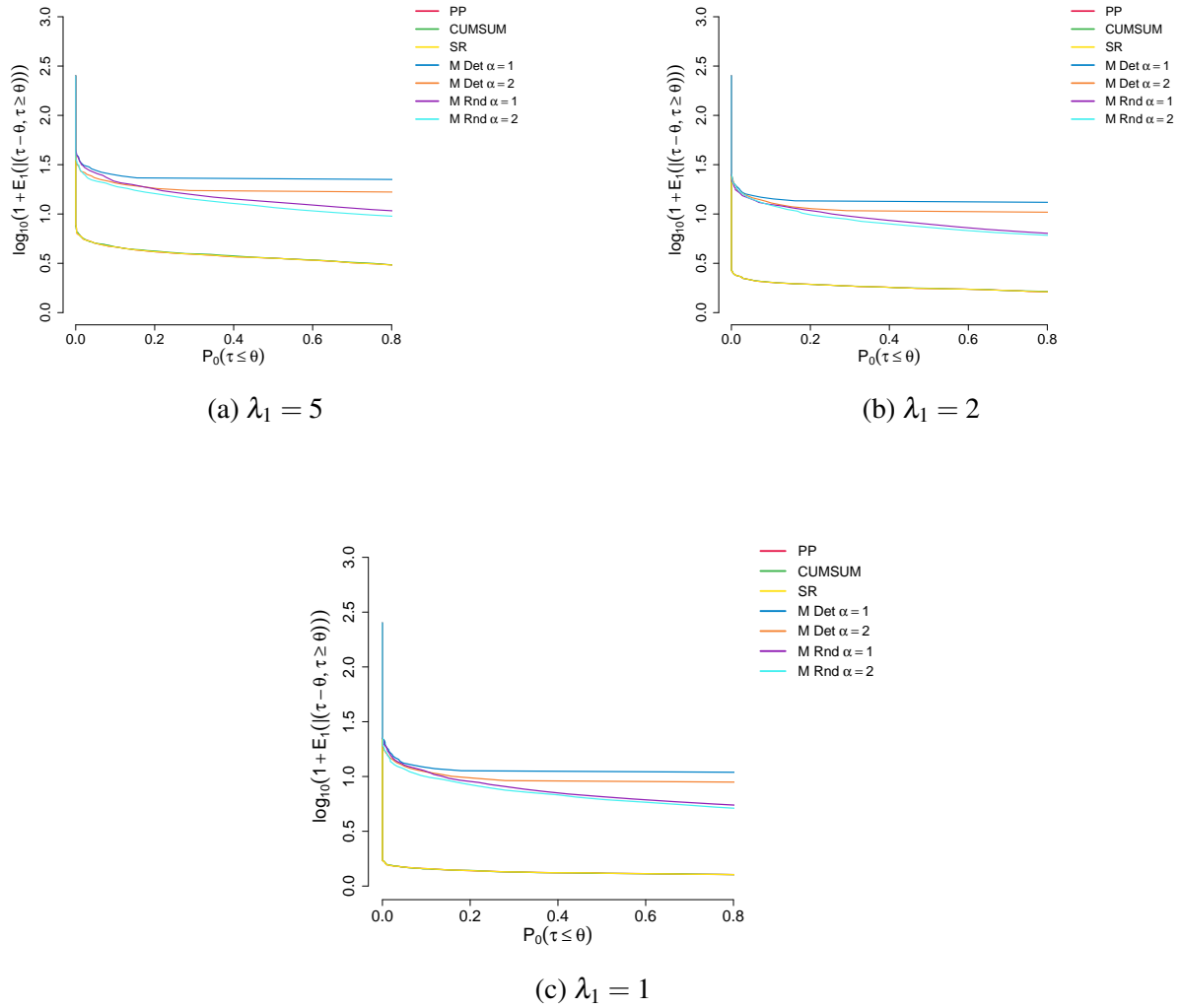


Figura 15 – Comparação dos métodos para distribuição Exponencial, com diferentes valores para λ_1 .

conhecidas na literatura, assim como uma medida de não conformidade baseada na distância energia (*e – distance*), introduzida por (SZÉKELY; RIZZO, 2004a).

No caso univariado, as medidas de não conformidade α^1 e α^2 foram definidas, respectivamente, como a menor distância absoluta e a média das distâncias absolutas entre uma nova observação e o conjunto de observações anteriores.

Para estender essas medidas ao cenário multivariado, foram realizadas adaptações naturais baseadas na distância euclidiana. Para cada nova observação $z_i \in \mathbb{R}^d$, foram calculadas:

- α^1 : a menor distância euclidiana entre z_i e os demais pontos observados até o instante n , isto é,

$$\alpha^1(z_i) = \min_{j \neq i} \sqrt{\sum_{k=1}^d (z_{i,k} - z_{j,k})^2},$$

em que $z_{i,k}$ representa a k -ésima coordenada de z_i ;

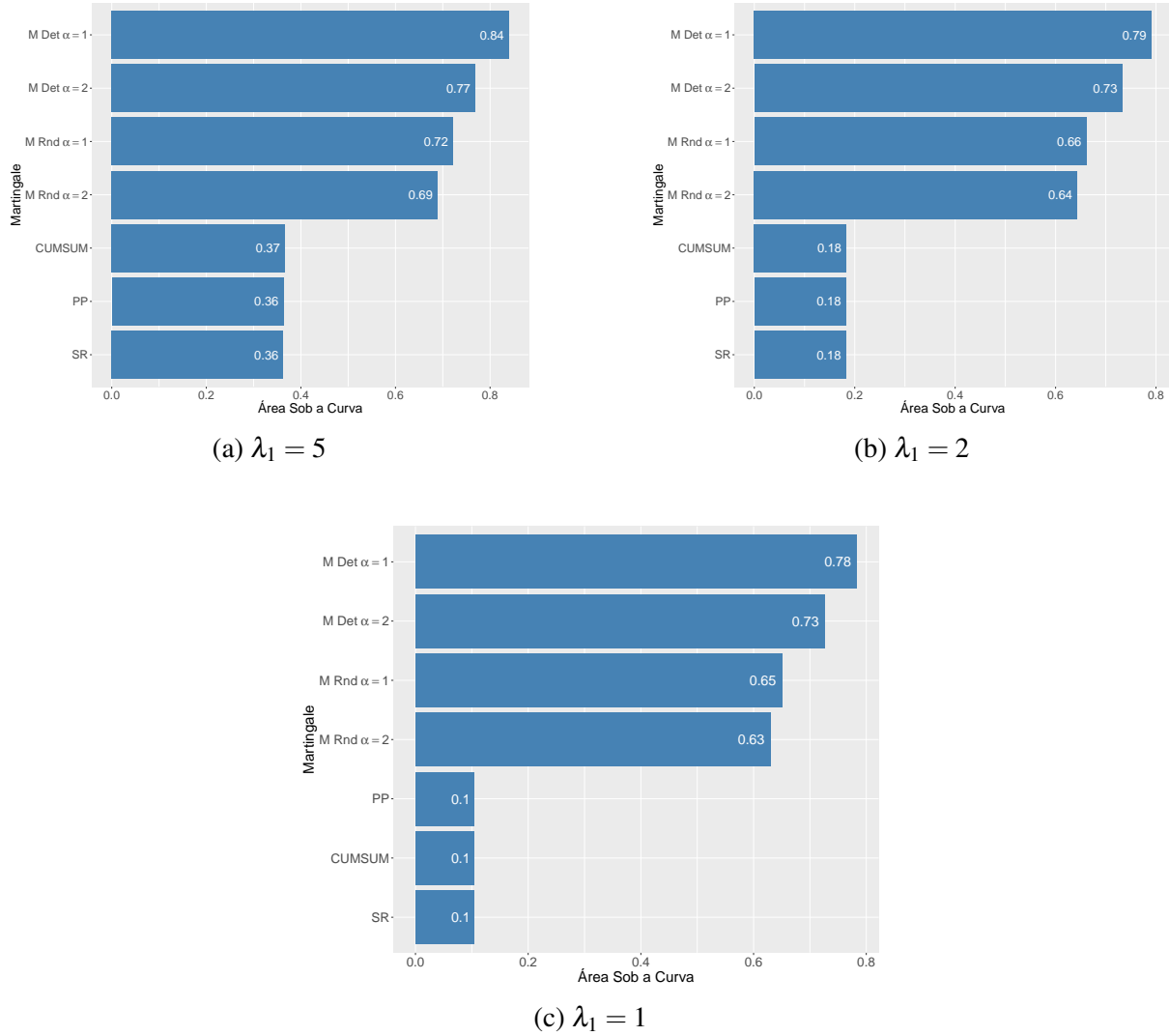


Figura 16 – AUC para distribuição Exponencial, com diferentes valores para λ_1 .

- α^2 : a média das distâncias euclidianas entre z_i e todos os demais pontos, dada por

$$\alpha^2(z_i) = \frac{1}{n-1} \sum_{j \neq i} \sqrt{\sum_{k=1}^d (z_{i,k} - z_{j,k})^2}.$$

Essas adaptações mantêm o espírito das definições originais: α^1 mede a similaridade um a um, pegando a distância mínima, enquanto α^2 mede uma noção global de dispersão em relação ao conjunto de dados observado até o instante atual.

Dessa forma, a transição do caso univariado para o multivariado foi realizada de maneira consistente, utilizando distâncias naturais no espaço \mathbb{R}^d .

4.3.1 Experimentos Multivariados de dimensão $d = 2$

A Tabela 7 apresenta as configurações que foram testadas para o caso multivariado com $d = 2$. Ou seja, além de serem testadas mudanças para os vetores de média μ (mantendo as

Tabela 7 – Configurações do experimento multivariado $d = 2$

Configuração	Parâmetros antes da mudança	Parâmetros após a mudança
Variação de μ	$\mu_0 = (0, 0), \quad \Sigma_0 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$	$\mu_1 = (1, 1), \quad \Sigma_1 = \Sigma_0$ $\mu_1 = (1.5, 1.5), \quad \Sigma_1 = \Sigma_0$ $\mu_1 = (2, 2), \quad \Sigma_1 = \Sigma_0$
Variação de Σ	$\mu_0 = (0, 0), \quad \Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\mu_1 = \mu_0, \Sigma_1 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$ $\mu_1 = \mu_0, \Sigma_1 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$

matrizes de variância e de covariância Σ_0 e Σ_1 iguais), também são realizadas simulações em que os vetores de médias são fixados e as matrizes de variância e covariância alteradas.

Na extensão para o cenário multivariado com dimensão $d = 2$, os resultados obtidos permaneceram amplamente coerentes com aqueles observados no contexto univariado.

Inicialmente, nos experimentos envolvendo variações apenas no vetor de médias μ_1 , com matriz de covariância constante ($\Sigma_0 = \Sigma_1$), verifica-se que os métodos clássicos (CUSUM, S-R, PP) novamente superaram os *conformal* martingales em termos de tempo médio de detecção, como ilustrado nas Figuras 17 e 18. Esse padrão se manteve independentemente da magnitude da mudança no vetor de médias, indicando a robustez dos métodos clássicos quando a especificação de \mathcal{P}^1 é exata.

Ao se analisar a variação da matriz de covariância (Σ_1) mantendo vetores de médias fixos ($\mu_0 = \mu_1$), representada nas Figuras 19 e 20, observa-se comportamento similar: os métodos clássicos mantêm superioridade em tempo de detecção, porém neste caso os *conformal* martingales apresentam resultados bem inferiores, indicando que os métodos não obtiveram sucesso ao detectar as mudanças na matriz de variância e covariância.

Importante destacar que, como no caso univariado, o desempenho dos martingales foi superior com o uso da medida de conformidade α^2 no experimento em que houve variação no vetor μ , que proporcionou tempos de detecção menores em comparação com α^1 .

De modo geral, os resultados multivariados reforçam as principais observações feitas no contexto univariado. Métodos clássicos são altamente eficazes quando a mudança no modelo é corretamente especificada e os martingales, embora menos competitivos em rapidez de detecção, oferecem certa robustez importante em contextos nos quais a especificação completa das distribuições não é feita.

Em suma, a adaptação para o espaço multivariado ($d = 2$) manteve o padrão de van-

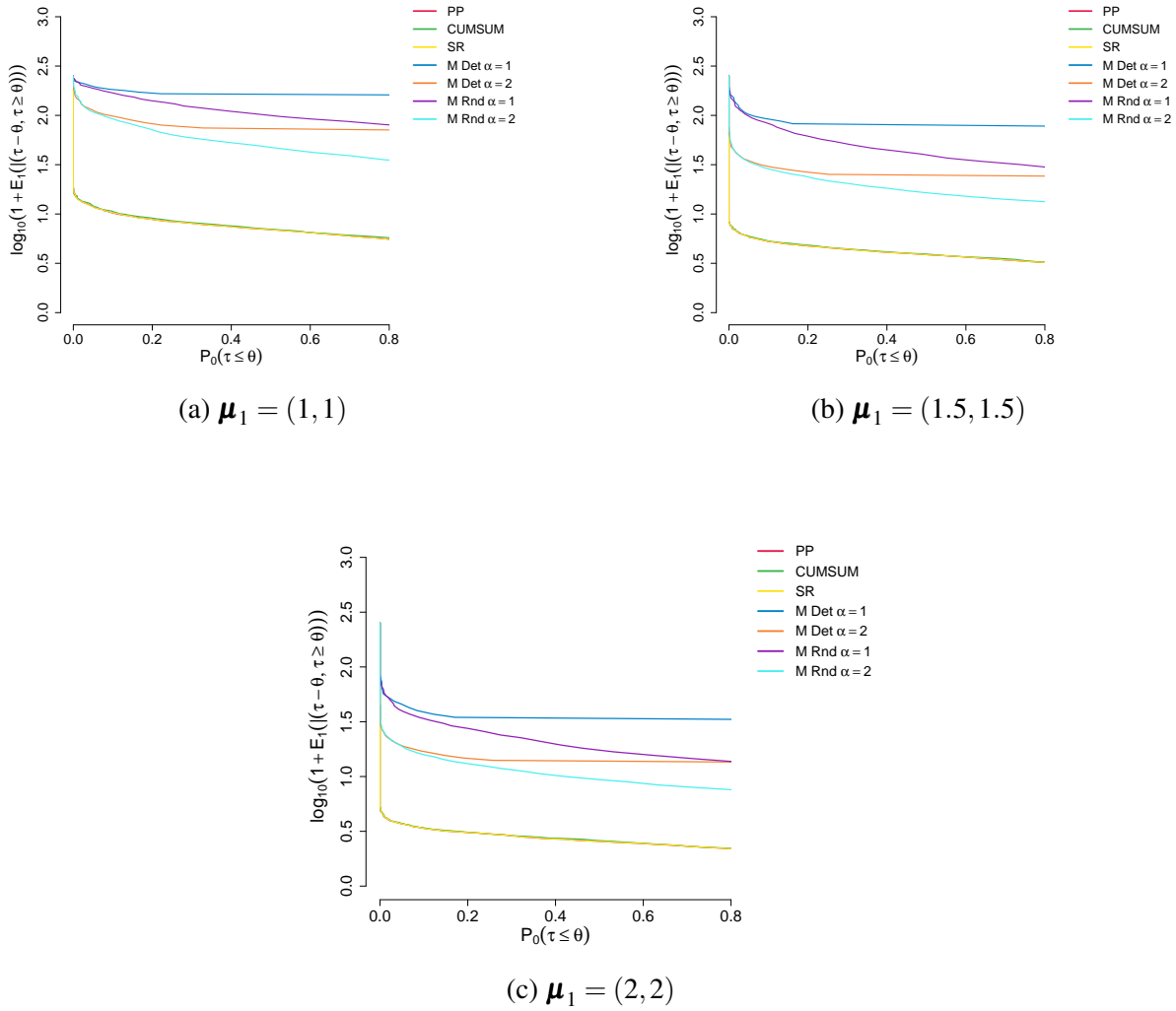


Figura 17 – Comparação dos métodos para distribuição Normal bivariada, com diferentes valores para o vetor de médias μ_1 , e matrizes de covariância $\Sigma_0 = \Sigma_1$ idênticas.

tagens e desvantagens já delineado, fortalecendo a compreensão comparativa entre métodos paramétricos clássicos e abordagens baseadas em *conformal* martingales.

4.3.2 Experimentos Multivariados de dimensão $d = 10$

Nesta seção, expandem-se as análises para o caso de dados multivariados com dimensão $d = 10$. Diferentemente dos experimentos com $d = 2$, os testes para $d = 10$ foram conduzidos em duas configurações distintas, com o intuito de avaliar a sensibilidade dos métodos clássicos e dos métodos baseados em *conformal* martingales à mudança na estrutura de variância e covariância.

Como mostra a Tabela 8, o vetor de médias permaneceu fixo em ambas as configurações, de modo a isolar o efeito das alterações na estrutura de covariância. Assim, o foco central recai sobre a capacidade dos métodos em detectar mudanças no padrão de correlação e na variabilidade dos dados.

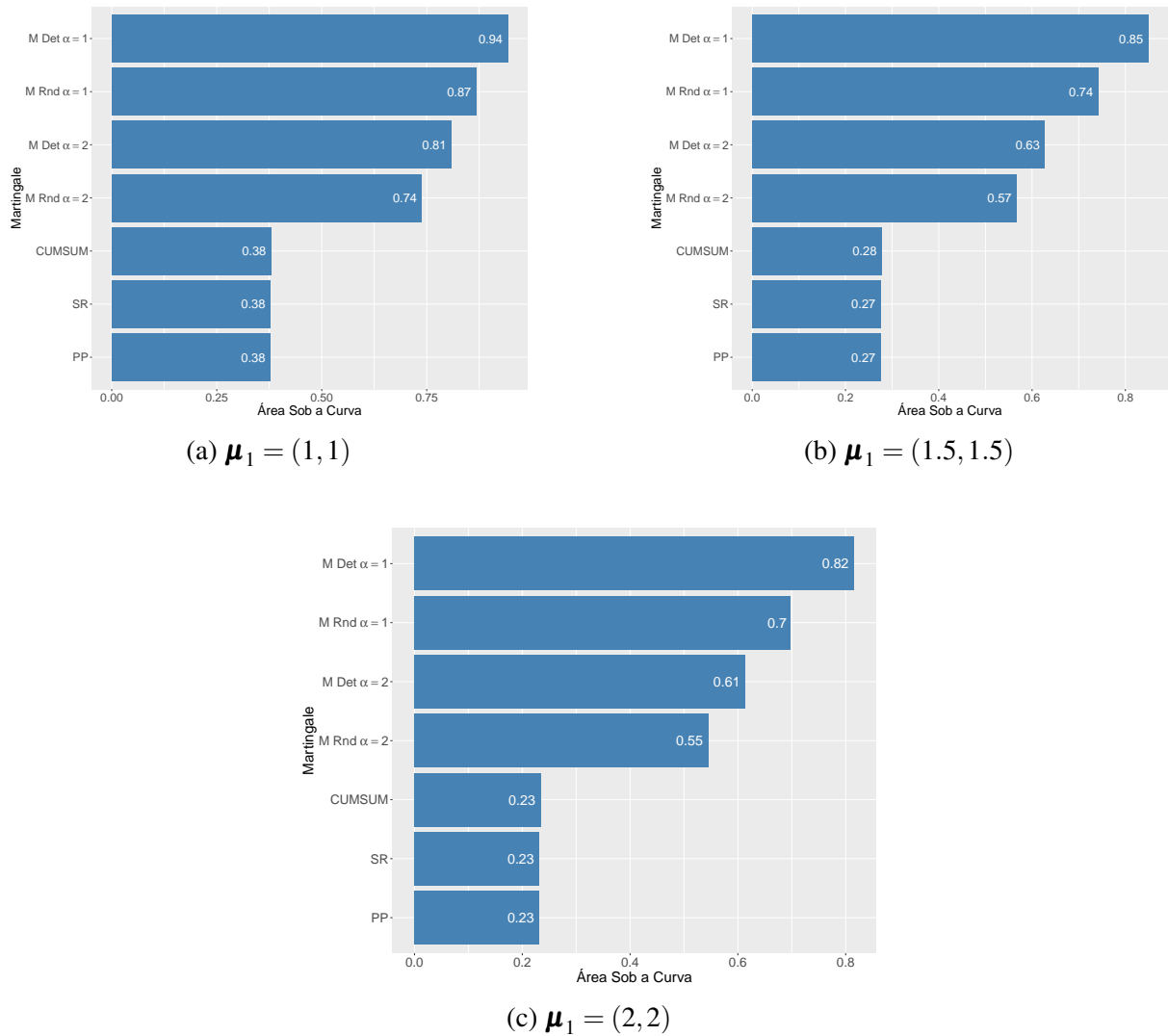


Figura 18 – AUC para distribuição Normal bivariada, com diferentes valores para o vetor de médias μ_1 , e matrizes de covariância idênticas $\Sigma_0 = \Sigma_1$.

Além das medidas de não conformidade α^1 e α^2 utilizadas nos experimentos com $d = 2$, foram incorporadas novas medidas, fundamentadas em conceitos de profundidade estatística e na distância de energia, com o objetivo de captar de maneira mais abrangente as discrepâncias entre as distribuições antes e após a mudança. Sendo assim, podemos adotar uma abordagem para a construção da medida de não conformidade utilizando diferentes medidas de profundidade estatística, que refletem o quão central uma observação está em relação a uma nuvem de dados multivariada. Neste caso, consideramos uma função de profundidade $\text{Depth}(\cdot, \mathcal{D}_{<i})$, que associa a cada nova observação $\mathbf{x}_i \in \mathbb{R}^d$ um valor de profundidade em relação ao conjunto de observações passadas $\mathcal{D}_{<i} = \{\mathbf{x}_1, \dots, \mathbf{x}_{i-1}\}$.

A medida de não conformidade α_i é então definida como a inversa regularizada do valor

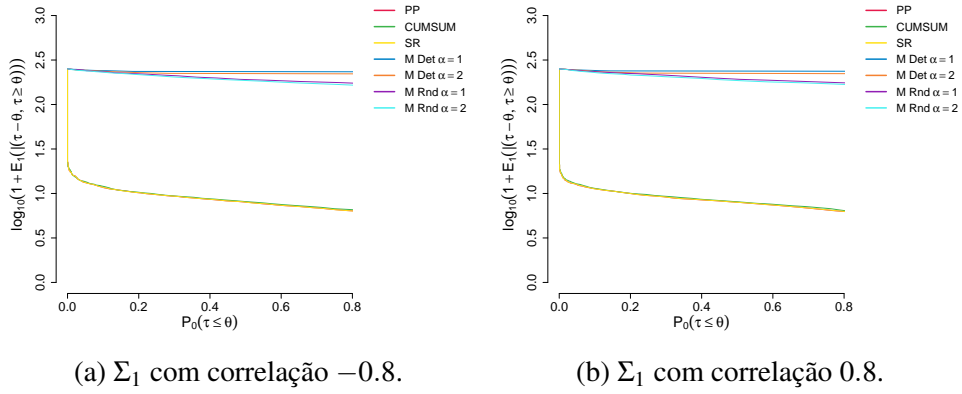


Figura 19 – Comparação dos métodos para distribuição Normal bivariada, com diferentes valores para $\Sigma_1 e (\mu_0 = \mu_1$ idênticos).

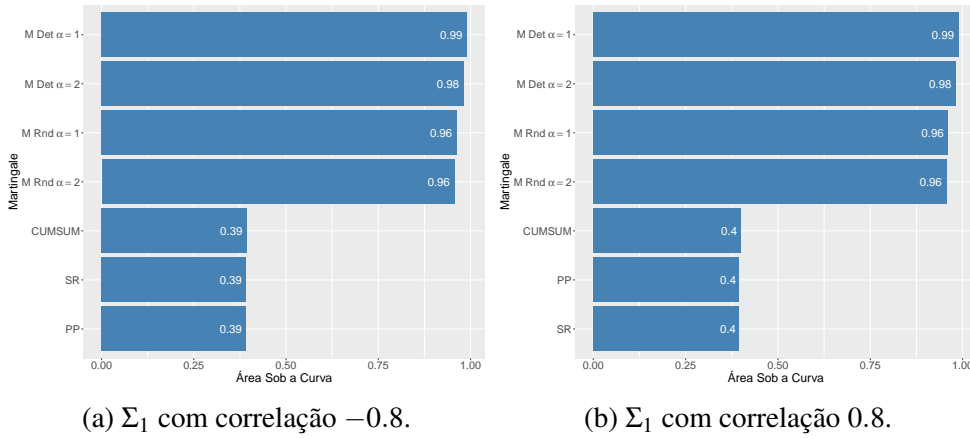


Figura 20 – AUC para Normal Bivariada com diferentes valores para $\Sigma_1 e (\mu_0 = \mu_1$ idênticos).

Tabela 8 – Configurações do experimento multivariado $d = 10$.

Configuração	Parâmetros antes da mudança	Parâmetros após a mudança
Varição de Σ (Config. 1)	$\mu_0 = \mu_1 = (0, \dots, 0) \in \mathbb{R}^{10}, \Sigma_0 = I_{10}$	$\Sigma_1 = \begin{pmatrix} 5 & 0.8 & \dots & 0.8 \\ 0.8 & 5 & \dots & 0.8 \\ \vdots & \vdots & \ddots & \vdots \\ 0.8 & 0.8 & \dots & 5 \end{pmatrix}$
Varição de Σ (Config. 2)	$\mu_0 = \mu_1 = (0, \dots, 0) \in \mathbb{R}^{10}, \Sigma_0 = I_{10}$	$\Sigma_1 = \begin{pmatrix} 5 & -0.1 & \dots & -0.1 \\ -0.1 & 5 & \dots & -0.1 \\ \vdots & \vdots & \ddots & \vdots \\ -0.1 & -0.1 & \dots & 5 \end{pmatrix}$

de profundidade:

$$\alpha_i := \frac{1}{\text{Depth}(\mathbf{x}_i, \mathcal{D}_{<i}) + \varepsilon}, \quad (4.1)$$

em que $\varepsilon = 10^{-6}$ é um termo de regularização adicionado para evitar instabilidades numéricas quando a profundidade for próxima de zero.

Essa formulação transforma medidas de centralidade em distâncias, atribuindo maiores

valores de não conformidade a observações mais periféricas (isto é, menos centrais em relação ao conjunto anterior). A função $\text{Depth}(\cdot, \cdot)$ pode ser implementada por diferentes métodos clássicos, neste trabalho utilizamos as seguintes medidas de profundidade:

- **Profundidade Espacial (L1)** (VARDI; ZHANG, 2000)

Para uma distribuição F em \mathbb{R}^d ,

$$D_{L1}(x; F) = \frac{1}{1 + \|S(x; F)\|}, \quad S(x; F) = \mathbb{E}_{X \sim F} \left[\frac{X-x}{\|X-x\|} \right].$$

Na prática, com a amostra $\{X_i\}_{i=1}^n$ (função depth.L1 do dalph no R),

$$\widehat{D}_{L1}(x) = \frac{1}{1 + \left\| \frac{1}{n} \sum_{i=1}^n \frac{X_i - x}{\|X_i - x\|} \right\|}.$$

Intuição: quanto mais próximo x estiver do “centro” L1 (mediana espacial), maior D_{L1} .

- **Profundidade L2** (MOSLER, 2013; KOSIOROWSKI; MOSLER, 2018)

$$D_{L2}(x; F) = \frac{1}{1 + \mathbb{E}_{X \sim F} \|X - x\|^2} \rightarrow \widehat{D}_{L2}(x) = \frac{1}{1 + \frac{1}{n} \sum_{i=1}^n \|X_i - x\|^2} \quad (\text{depth.L2 no R}).$$

Observação: utiliza distâncias quadráticas; é mais sensível a outliers que D_{L1} , porém simples de computar.

- **Profundidade por Potencial** (MOSLER, 2013; KOSIOROWSKI; MOSLER, 2018)

Para um parâmetro $\beta > 0$,

$$D_{\text{pot}}(x; F) = \frac{1}{1 + \mathbb{E}_{X \sim F} \|X - x\|^\beta} \rightarrow \widehat{D}_{\text{pot}}(x) = \frac{1}{1 + \frac{1}{n} \sum_{i=1}^n \|X_i - x\|^\beta} \quad (\text{depth.potential no R}).$$

Intuição: pontos densos (centrais) geram valores elevados; pontos periféricos recebem profundidade baixa.

- **Distância de Energia (e-distance)** (SZÉKELY; RIZZO, 2004b; RIZZO; SZÉKELY, 2023)

Para duas distribuições F e G ,

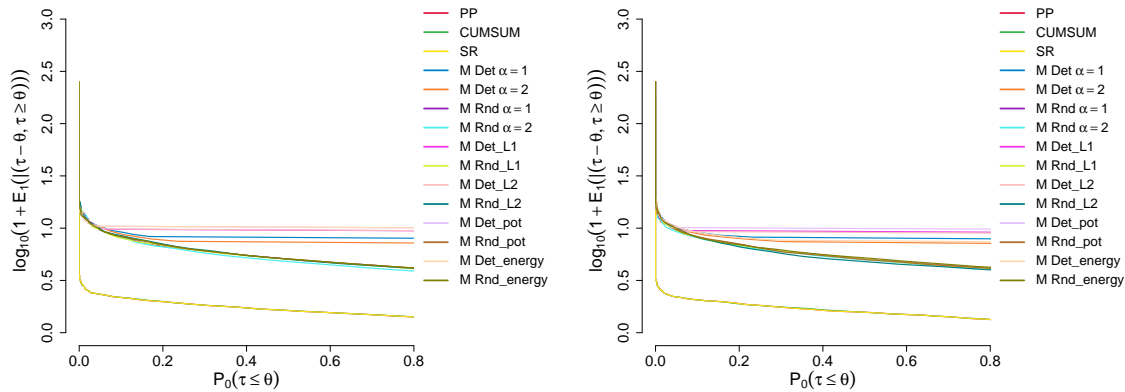
$$\mathcal{E}(F, G) = 2 \mathbb{E} \|X - Y\| - \mathbb{E} \|X - X'\| - \mathbb{E} \|Y - Y'\|,$$

em que $X, X' \sim F$ e $Y, Y' \sim G$ (independentes). No teste de conformidade de uma nova observação x contra a amostra $\{X_i\}_{i=1}^n$ (função edist do pacote energy no R), empregamos a estatística de não-conformidade

$$N_e(x) = \frac{1}{n} \sum_{i=1}^n \|X_i - x\| - \frac{1}{2n^2} \sum_{i,j=1}^n \|X_i - X_j\|.$$

Propriedade: sensível a diferenças de escala, variância e correlação – altamente eficaz em alta dimensão.

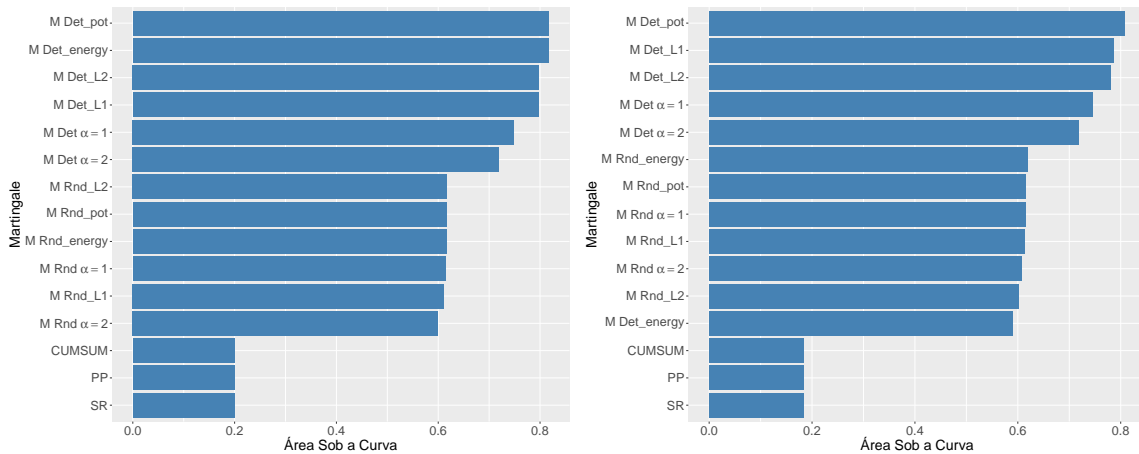
Em cada uma das configurações, os métodos clássicos (CUSUM, Shiryaev-Roberts e PP) e os métodos baseados em *conformal* martingales foram aplicados para detectar a mudança na estrutura de covariância. As medidas de não conformidade tradicionais α^1 e α^2 foram comparadas com as novas medidas baseadas em profundidade (Profundidade Espacial (L1), Profundidade L2, Potencial e *e-distance*), permitindo uma análise mais ampla do desempenho sob cenários de alta dimensão. Assim como anteriormente, a avaliação dos métodos foi conduzida por meio da análise da probabilidade de falso positivo e do tempo médio para detecção, possibilitando uma comparação direta entre as diferentes abordagens.



(a) Configuração 1 da Tabela 8.

(b) Configuração 2 da Tabela 8.

Figura 21 – Comparação dos métodos para distribuição Normal Multivariada com $d = 10$, à esquerda, Configuração 1 da Tabela 8, à direita, Configuração 2.



(a) Curvas taxa de alarme falso por tempo de detecção para Configuração 1 da tabela 8.

(b) Curvas taxa de alarme falso por tempo de detecção para Configuração 2 da tabela 8.

Figura 22 – AUC para distribuição Normal Multivariada com $d = 10$, à esquerda, Configuração 1 da Tabela 8, à direita, Configuração 2.

Os experimentos realizados para dados multivariados de dimensão $d = 10$, ilustrados nas Figuras 21 e 22, revelam comportamentos semelhantes aos observados para $d = 2$, mas destacam nuances importantes na avaliação das medidas de não conformidade.

Assim como nos experimentos anteriores, os métodos clássicos mantêm desempenho superior, apresentando menores tempos médios de detecção para as mesmas taxas de falso alarme. No entanto, observou-se que, entre os *conformal* martingales, a escolha da medida de não conformidade torna-se ainda mais crítica em alta dimensão.

Na Configuração 1 da Tabela 8, os martingales randomizados com medidas α^2 e profundidade L1 apresentaram as menores AUCs, indicando melhor desempenho. Já na Configuração 2, as medidas baseadas em distância de energia (*energy*) e profundidade L2 se destacaram.

Analisando os gráficos de AUC, nota-se um padrão interessante: há um grupo inicial de martingales, composto pelos martingales baseados em *transducers* randomizados, que apresentam desempenhos visivelmente melhores que os martingales determinísticos. Esse comportamento evidencia que, embora existam pequenas variações entre as melhores medidas, escolher o tipo de *transducer* implica em diferenças significativas nos resultados finais.

Entre as novas medidas testadas, destaca-se que a profundidade por potencial (*pot*) apresentou, de modo geral, desempenho inferior às demais, com AUCs consistentemente mais elevadas em ambas as configurações, principalmente com o *transducer determinístico*.

Ainda que os martingales continuem apresentando maior atraso médio de detecção em comparação com os métodos clássicos, os resultados obtidos reforçam que, com uma escolha adequada da medida de não conformidade, é possível obter desempenhos competitivos mesmo em cenários de alta dimensão e sob mudanças sutis na estrutura de covariância.

De maneira geral, os experimentos com $d = 10$ reiteram a vantagem dos métodos clássicos quando as distribuições são corretamente especificadas, mas também demonstram a flexibilidade dos *conformal* martingales, especialmente quando combinados com medidas de conformidade adequadas.

Em conclusão, os resultados para $d = 10$ reforçam que a escolha da medida de não conformidade torna-se ainda mais crucial em cenários de alta dimensão. Embora os métodos clássicos continuem liderando sob hipótese de especificação correta, certas combinações de martingales determinísticos com medidas baseadas em profundidade L1, distância de energia ou α^2 conseguem se aproximar competitivamente, especialmente em estruturas complexas de covariância.

4.4 Especificação Incorreta em Métodos Clássicos

Como já mencionado algumas vezes anteriormente, uma desvantagem significativa das abordagens clássicas é a necessidade de conhecimento completo das distribuições \mathcal{P}^0 e \mathcal{P}^1 , o que raramente é viável em problemas práticos. Assim, esses métodos não são adequados para cenários cujas distribuições antes e depois da mudança são desconhecidas.

Para demonstrar que o bom desempenho dos métodos clássicos depende fortemente de uma especificação correta das distribuições pré e pós mudança, foram realizados dois experimentos adicionais, da mesma forma que os experimentos anteriores, com 1000 réplicas de vetores de tamanho 500, um experimento com dados gerados sob uma distribuição Normal univariada e outro experimento de uma Normal multivariada com $d = 10$, em que as especificações estão descritas a seguir.

Configuração com especificação incorreta dos parâmetros

Univariado (Normal)

- Parâmetros reais (geração dos dados): $\mu_0 = 0$, $\mu_1 = 2$, $\sigma_0 = \sigma_1 = 1$
- Parâmetros assumidos pelos métodos clássicos: $\mu_0 = 0$, $\mu_1 = 10$, $\sigma_0 = \sigma_1 = 1$

Multivariado (Normal, $d = 10$)

- Parâmetros reais (geração dos dados):

$$\mu_0 = \mu_1 = 0, \quad \Sigma_0 = \begin{bmatrix} 1 & 0 & \dots \\ 0 & \ddots & \\ \vdots & & 1 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 1 & 0.1 & \dots \\ 0.1 & \ddots & \\ \vdots & & 1 \end{bmatrix}$$

- Parâmetros assumidos pelos métodos clássicos:

$$\mu_0 = \mu_1 = 0, \quad \Sigma_0 = \begin{bmatrix} 1 & 0 & \dots \\ 0 & \ddots & \\ \vdots & & 1 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 1 & 0.8 & \dots \\ 0.8 & \ddots & \\ \vdots & & 1 \end{bmatrix}$$

Os resultados obtidos nos experimentos, apresentados nas Figuras 23 e 24, evidenciam de forma clara a alta sensibilidade dos métodos clássicos à correta especificação da distribuição após a mudança (H_1). Quando essa especificação é incorreta, como neste caso, observa-se uma degradação no desempenho dos métodos tradicionais (CUSUM, Shiryaev-Roberts e PP).

No cenário **univariado** (Figura 23), percebe-se que os métodos clássicos apresentaram áreas abaixo das curvas (AUCs) significativamente maiores, indicando tempos médios de detecção piores. Em contraste, os *conformal* martingales, que não dependem da parametrização explícita, mantiveram seu padrão de desempenho, superando inclusive os métodos clássicos nas métricas analisadas nos cenários testados.

De maneira análoga, no experimento **multivariado** com $d = 10$ (Figura 24), o comportamento se repete: os métodos clássicos, ao operarem sob informações incorretas da matriz

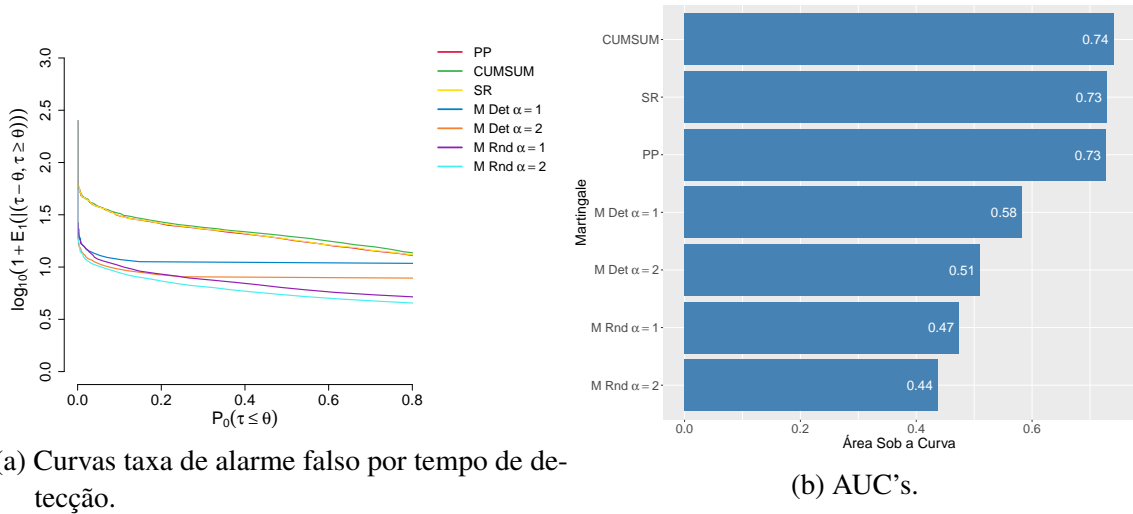


Figura 23 – Resultados do experimento adotando valores errados para o caso **univariado**, à esquerda, as curvas da relação de taxa de alarme falsos por tempo de detecção e à direita, a área abaixo da curva.

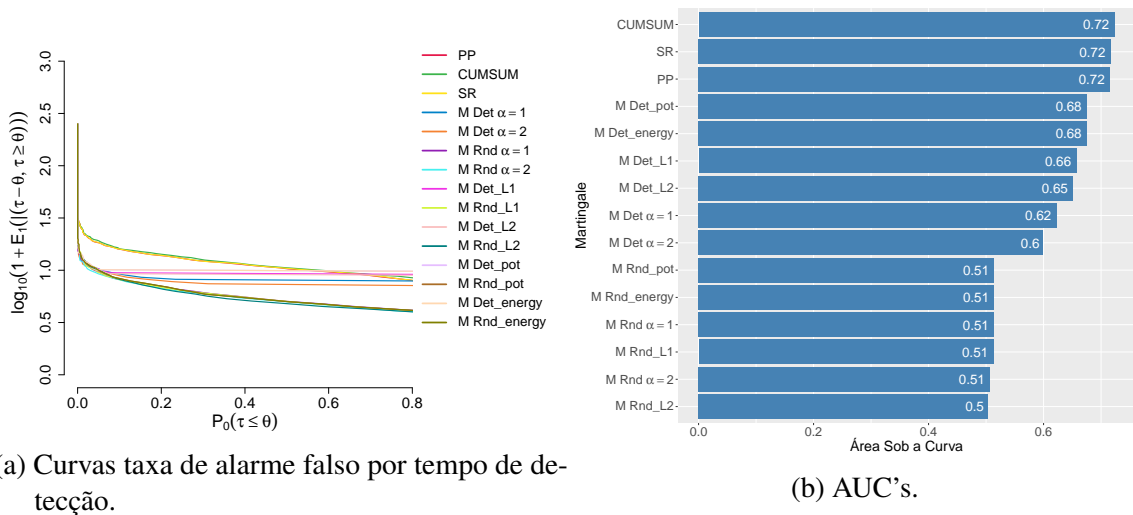


Figura 24 – Resultados do experimento adotando valores errados para o caso **multivariado**, à esquerda, as curvas da relação de taxa de alarme falsos por tempo de detecção e à direita, a área abaixo da curva.

de covariância Σ_1 , tiveram perdas substanciais na eficácia da detecção. Em vários cenários, os martingales, mesmo com atraso característico de sua natureza mais geral, foram capazes de superar os métodos tradicionais, apresentando AUCs inferiores.

Esses resultados reforçam, com evidência empírica, uma das principais motivações para o uso de métodos baseados em *conformal* martingales: a falta de conhecimento das distribuições dos modelos geradores dos dados. Em ambientes práticos, nos quais conhecer a verdadeira distribuição pós-mudança é desafiador ou mesmo inviável, métodos que dispensam tal conhecimento tornam-se uma escolha mais segura, ainda que à custa de tempos médios de detecção ligeiramente superiores em situações ideais.

Portanto, a escolha entre métodos clássicos e *conformal* martingales deve ser pautada principalmente pela confiança que se tem sobre o conhecimento das distribuições. Em cenários com elevada incerteza sobre os dados, os *conformal* martingales se destacam como uma alternativa promissora e mais resiliente.

ESTUDOS EM LOTES DE DADOS SEQUENCIAIS

A detecção automática de mudanças de distribuição (*dataset shift*) é um tema central em aprendizagem de máquina porque o desempenho de modelos preditivos costuma degradar rapidamente quando as suposições de invariância de distribuição são violadas (QUIÑONERO-CANDELA *et al.*, 2009; MORENO-TORRES *et al.*, 2012a). Em aplicações em que os dados chegam em blocos temporais — por exemplo, atualizações mensais de transações financeiras, coletas periódicas de sensores industriais ou rodadas sucessivas de exames laboratoriais — torna-se natural tratar cada bloco como uma unidade (*lote*) e investigar, a cada nova chegada, se a distribuição geradora mudou em relação aos lotes anteriores.

Existem pelo menos quatro formas clássicas de *shift*:

1. **Covariate shift** (\mathbb{P}_X muda, $\mathbb{P}_{Y|X}$ permanece): É típico em sistemas de recomendação ou pontuação de crédito, nos quais o perfil dos usuários (covariáveis) se altera com o tempo, mas a relação entre perfis e inadimplência se mantém aproximadamente estável (SHIMODAIRA, 2000);
2. **Label shift** (\mathbb{P}_Y muda, $\mathbb{P}_{X|Y}$ permanece): Surge em cenários de detecção de doenças raras, em que a prevalência da classe positiva varia entre hospitais ou ao longo de surtos, mas a distribuição de características condicionada ao diagnóstico é constante (LIPTON; WANG; SMOLA, 2018);
3. **Concept shift tipo I** ($\mathbb{P}_{X|Y}$ muda): Também chamado *feature-conditional shift*, aparece em filtragem de *spam*. À medida que *spammers* adaptam seu vocabulário, a distribuição das palavras dadas as classes “spam” ou “ham” evolui, mesmo que a proporção de *spams* permaneça a mesma (GAMA *et al.*, 2014);

4. **Total dataset shift** (mudança em $\mathbb{P}_{X,Y}$): Um exemplo prático ocorre em sensores químicos que sofrem *drift* devido ao envelhecimento. Nesse caso, tanto as leituras dos sensores (covariáveis) quanto as concentrações reais das substâncias medidas (rótulos) podem sofrer mudanças ao longo do tempo (ALIPPI; ROVERI, 2008);

Caracterizar e detectar cada tipo de mudança é essencial não só para acionar alarmes ou re-treinar modelos, mas também para diagnosticar a origem da degradação e escolher estratégias de adaptação adequadas (WIDMER; KUBAT, 1996). Este capítulo formaliza o problema em termos de lotes sequenciais de dados e introduz uma abordagem baseada em divergência de Kullback–Leibler e *conformal* martingales para monitorar, em tempo real, esses diferentes tipos de *shift*.

5.1 Definição do problema

Considere observações $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2), \dots, Z_n = (X_n, Y_n), \dots$ observações com suporte em \mathbb{R}^{p+1} em que supomos que estas observações estejam agrupadas em *datasets* com mesma distribuição, isto é,

$$\begin{aligned} Z_1, \dots, Z_{b_1} &\stackrel{i.i.d.}{\sim} \mathbb{P}_{X,Y}^1, \\ Z_{b_1+1}, \dots, Z_{b_1+b_2} &\stackrel{i.i.d.}{\sim} \mathbb{P}_{X,Y}^2, \\ &\dots \\ Z_{b_1+\dots+b_{i-1}+1}, \dots, Z_{b_1+\dots+b_i} &\stackrel{i.i.d.}{\sim} \mathbb{P}_{X,Y}^i, \end{aligned}$$

para $i \geq 1$.

O problema a ser resolvido consiste em avaliar as seguintes hipóteses

$$H_{0,T}^i : \mathbb{P}_{X,Y}^1 = \dots = \mathbb{P}_{X,Y}^i \quad (5.1)$$

$$H_{0,C}^i : \mathbb{P}_X^1 = \dots = \mathbb{P}_X^i \quad (5.2)$$

$$H_{0,L}^i : \mathbb{P}_Y^1 = \dots = \mathbb{P}_Y^i \quad (5.3)$$

$$H_{0,C1}^i : \mathbb{P}_{X|Y}^1 = \dots = \mathbb{P}_{X|Y}^i \quad (5.4)$$

$$H_{0,C2}^i : \mathbb{P}_{Y|X}^1 = \dots = \mathbb{P}_{Y|X}^i \quad (5.5)$$

com base nas observações $Z_1, \dots, Z_{b_1+\dots+b_i}$, $i \geq 1$.

A hipótese nula (5.1) indica que não há diferença entre os conjuntos de dados no que diz respeito à distribuição total; a hipótese (5.2) indica que não há diferença entre as distribuições quanto a distribuição das covariáveis; (5.3) para a distribuição do *label* (variável resposta em um contexto de classificação); (5.4) para a distribuição condicional de $(X|Y)$ e (5.5) para distribuição condicional de $(Y|X)$.

Sendo assim, organizamos as observações $Z_i = (X_i, Y_i)$ em lotes ou subconjuntos de dados consecutivos, os quais denotamos por $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_N$, em que cada lote \mathcal{D}_i é definido como

$$\mathcal{D}_i = \{Z_{b_0+b_1+\dots+b_{i-1}+1}, \dots, Z_{b_0+b_1+\dots+b_i}\},$$

com b_i sendo o número de observações do i -ésimo lote e $b_0 = 0$.

Com essa notação, cada lote \mathcal{D}_i contém b_i observações consecutivas, começando imediatamente após o término do lote anterior. Por exemplo, se todos os lotes tiverem tamanho fixo $b_i = 500$, então temos: $\mathcal{D}_1 = \{Z_1, \dots, Z_{500}\}$, $\mathcal{D}_2 = \{Z_{501}, \dots, Z_{1000}\}$, e assim sucessivamente.

Cada lote \mathcal{D}_i é composto por b_i observações independentes e identicamente distribuídas de acordo com uma distribuição $\mathbb{P}_{X,Y}^{(i)}$, conforme definido anteriormente. Dessa forma, o problema de detecção de mudança de distribuição entre observações individuais passa a ser modelado como um problema de detecção de mudança entre distribuições associadas a cada lote \mathcal{D}_i .

Neste contexto, os lotes \mathcal{D}_i passam a ser tratados como as unidades sequenciais de interesse, e a tarefa consiste em verificar, a cada nova observação \mathcal{D}_n , se há evidência de mudança de distribuição em relação aos lotes anteriores.

5.2 Kullback-Leibler

Para testar as hipóteses apresentadas anteriormente, primeiramente, vamos definir a divergência de *Kullback-Leibler*, utilizada para medir o quanto dois *datasets* divergem em distribuição entre si.

A divergência de *Kullback-Leibler* é definida formalmente como:

$$\text{KL}(\mathcal{P} \parallel \mathcal{Q}) := \int \log \left(\frac{d\mathcal{P}}{d\mathcal{Q}} \right) d\mathcal{P}$$

em que $\left(\frac{d\mathcal{P}}{d\mathcal{Q}} \right)$ é a derivativa de Radon–Nikodym (COVER; THOMAS, 2006), assumindo que \mathcal{Q} domina \mathcal{P} , que é a densidade de \mathcal{P} em respeito a \mathcal{Q} .

A medida *KL* é utilizada para detectar as diferenças entre dois conjuntos de dados seja qual tipo de diferença estamos tentando identificar, *label shift*, *concept shift* (tipo 1 e 2), *covariate shift* e *total dataset shift*, cuja a magnitude dos valores de *KL* é diretamente relacionada com a intensidade do *shift* a ser analisado. Sendo assim, considerando dois conjunto de dados $D^{(1)}$ e $D^{(2)}$, em que $\mathcal{P}^{(1)}$ é a distribuição de probabilidade do *dataset* $D^{(1)}$ e $\mathcal{P}^{(2)}$ do *dataset* $D^{(2)}$, para calcular a quantidade de *shift* entre os dois *datasets*, são definidas as seguintes quantidades teóricas:

$$\text{[Total Dataset Shift]} \text{KL}_{X,Y} =: (\mathcal{P}_{X,Y}^{(2)} \parallel \mathcal{P}_{X,Y}^{(1)});$$

$$\text{[Covariate Shift]} \text{ } KL_X =: (\mathcal{P}_X^{(2)} \parallel \mathcal{P}_X^{(1)})$$

$$\text{[Label Shift]} \text{ } KL_Y =: (\mathcal{P}_Y^{(2)} \parallel \mathcal{P}_Y^{(1)});$$

$$\text{[Concept Shift Tipo 1]} \text{ } KL_{X|Y} =: (\mathcal{P}_{X|Y}^{(2)} \parallel \mathcal{P}_{X|Y}^{(1)});$$

$$\text{[Concept Shift Tipo 2]} \text{ } KL_{Y|X} =: (\mathcal{P}_{Y|X}^{(2)} \parallel \mathcal{P}_{Y|X}^{(1)});$$

5.2.1 Estimativa da divergência de Kullback-Leibler

Para estimar as divergências $KL_{X,Y}$, KL_X e KL_Y , seguimos a abordagem proposta por (CRANMER; PAVEZ; LOUPPE, 2016) e (SUGIYAMA; SUZUKI; KANAMORI, 2012), utilizando um classificador probabilístico treinado para prever de qual distribuição uma amostra provém. Esse classificador é treinado no conjunto \mathcal{D}^{tr} formado por:

$$\mathcal{D} = \{(X_k, Y_k, Z_k)\}_{k=1}^n, \quad Z_k \in \{1, 2\},$$

em que Z_k indica se a amostra (X_k, Y_k) pertence à população 1 ou 2. O classificador retorna a probabilidade condicional $\widehat{\mathbb{P}}(Z = i \mid X, Y)$, usada para estimar a derivada de Radon–Nikodym. A estimativa da razão de densidade entre $\mathcal{P}^{(2)}$ e $\mathcal{P}^{(1)}$ para o caso de shift total, por exemplo, é dada por:

$$\frac{\widehat{d\mathcal{P}_{X,Y}^{(2)}}}{\widehat{d\mathcal{P}_{X,Y}^{(1)}}}(x, y) := \frac{n_{tr}^{(1)}}{n_{tr}^{(2)}} \cdot \frac{\widehat{\mathbb{P}}(Z = 2 \mid X = x, Y = y)}{\widehat{\mathbb{P}}(Z = 1 \mid X = x, Y = y)}.$$

A divergência é então calculada via média empírica sobre o conjunto de teste $\mathcal{D}_2^{te} = \{(X_k, Y_k, Z_k) \in \mathcal{D}^{te} \mid Z_k = 2\}$:

$$\widehat{KL}_{X,Y} := \frac{1}{|\mathcal{D}_2^{te}|} \sum_{(X_k, Y_k) \in \mathcal{D}_2^{te}} \log \left(\frac{\widehat{d\mathcal{P}_{X,Y}^{(2)}}}{\widehat{d\mathcal{P}_{X,Y}^{(1)}}}(X_k, Y_k) \right),$$

de maneira análoga para \widehat{KL}_X e \widehat{KL}_Y .

Contudo, para os casos condicionais ($KL_{X|Y}$, $KL_{Y|X}$), esse procedimento não é diretamente aplicável. Nestes casos, usamos a decomposição da divergência KL, conforme o Teorema 2.13 de (WU, 2022), que afirma:

$$KL_{X,Y} = KL_{X|Y} + KL_Y = KL_{Y|X} + KL_X.$$

permitindo estimar as divergências condicionais via:

$$\widehat{\text{KL}}_{X|Y} := \widehat{\text{KL}}_{X,Y} - \widehat{\text{KL}}_Y, \quad \widehat{\text{KL}}_{Y|X} := \widehat{\text{KL}}_{X,Y} - \widehat{\text{KL}}_X.$$

que é conhecida como regra da cadeia para divergência de Kullback–Leibler.

5.3 Definição dos experimentos

Após definirmos a divergência de Kullback–Leibler (KL) e a metodologia utilizada para estimá-la, podemos agora descrever como ela será utilizada como medida de não conformidade. Neste capítulo, são apresentadas comparações realizadas entre duas abordagens diferentes para calcular as medidas α_i , as quais serão posteriormente utilizadas como entradas (inputs) para os *conformal transducers*, definidos previamente nos Experimentos (2.3) e (2.5).

Considere uma sequência de conjuntos de dados $\{\mathcal{D}_i\}_{i=1}^N$, representando lotes (*datasets*) sequenciais de observações conforme definido na Seção 5.1. Para cada novo conjunto de dados \mathcal{D}_n , $n = 2, \dots, N$, calculamos a divergência KL entre o novo conjunto e cada um dos conjuntos anteriores. Essa divergência KL será estimada de acordo com o tipo de *dataset shift* que desejamos detectar (*total shift*, *label shift*, *covariate shift* ou *concept shift*).

Neste contexto, definimos duas medidas alternativas para representar a não conformidade, utilizando a mesma ideia de α_i^1 e α_i^2 , utilizadas nas Seções 3.1 e 3.2, porém substituindo a distância euclidiana pela medida KL:

- A medida α_i^1 , definida como a divergência KL mínima entre o conjunto \mathcal{D}_i e todos os outros conjuntos já observados até o conjunto atual \mathcal{D}_n :

$$\alpha_i^1 = \min_{1 \leq j \leq n, j \neq i} \left\{ \widehat{\text{KL}}(\mathcal{D}_i, \mathcal{D}_j) \right\}. \quad (5.6)$$

- A medida α_i^2 , definida como a divergência KL média entre o conjunto \mathcal{D}_i e todos os outros conjuntos já observados até o conjunto atual \mathcal{D}_n :

$$\alpha_i^2 = \frac{1}{n-1} \sum_{1 \leq j \leq n, j \neq i} \widehat{\text{KL}}(\mathcal{D}_i, \mathcal{D}_j). \quad (5.7)$$

Assim, para cada novo conjunto de dados inserido na sequência, as medidas de não conformidade α_i^1 e α_i^2 são recalculadas, representando, respectivamente, o grau de não conformidade do conjunto em relação ao conjunto mais similar (medida mínima) e ao conjunto médio dos dados anteriores (medida média). Estas medidas são então utilizadas diretamente no *conformal transducer* determinístico ou randomizado, seguindo a metodologia geral apresentada e utilizada nos capítulos anteriores, estendida aqui para o uso da divergência KL entre distribuições de lotes sequenciais.

5.4 Simulação

Uma vez definida a metodologia para calcular as medidas de não conformidade e gerar os valores p^c , foi realizado um experimento com dados simulados. Este experimento tem como objetivo avaliar o desempenho das duas medidas de não conformidade propostas (α^1 e α^2) utilizando a KL, bem como verificar a sensibilidade dos *conformal* martingales em detectar diferentes tipos de *dataset shift*.

5.4.1 Experimento

Neste experimento os dados simulados consistem em réplicas de uma sequência de 200 conjuntos de dados (unidades observacionais de interesse neste capítulo), cada um contendo 500 observações (número de linhas do *dataset*). Os primeiros 100 conjuntos são gerados a partir de uma distribuição inicial, enquanto os 100 conjuntos subsequentes são gerados a partir de uma nova distribuição (muito parecido com os experimentos anteriores, porém com apenas **250 réplicas** dado a limitações computacionais), simulando explicitamente os seguintes tipos de *dataset shift*: *label shift*, *total dataset shift*, *covariate shift* e *concept shift tipo 1*. Utilizamos uma estratégia para geração de dados similar à utilizada por (MAIA POLO *et al.*, 2023), porém adaptando ao nosso contexto, no qual são gerados vários *datasets*. As distribuições utilizadas para gerar os dados são dadas por:

$$\mathcal{P}_Y^{(1)} = \text{Ber}\left(\frac{1}{2}\right), \quad \mathcal{P}_Y^{(2)} = \text{Ber}\left(\frac{1}{2} + \delta\right),$$

$$\mathcal{P}_{X|Y}^{(1)} = \mathcal{N}(Y \cdot \mathbf{1}_d, \mathbf{I}_d), \quad \mathcal{P}_{X|Y}^{(2)} = \mathcal{N}((Y + \gamma) \cdot \mathbf{1}_d, \mathbf{I}_d),$$

em que $\text{Ber}(p)$ denota a distribuição de Bernoulli com probabilidade p , $\mathbf{1}_d$ é o vetor de uns com dimensão $d = 10$, e $\mathcal{N}(\mu, \Sigma)$ é a distribuição normal multivariada com vetor médio μ e matriz de covariância Σ .

Neste contexto, δ controla o nível de *label shift*, afetando diretamente a divergência marginal KL_Y , enquanto γ controla o nível de *concept shift tipo 1* e, consequentemente, *covariate shift*, afetando diretamente a divergência condicional $\text{KL}_{X|Y}$ e KL_X .

Nos nossos experimentos, fixamos inicialmente os valores $\delta = 0$ e $\gamma = 0$, de modo que os 100 primeiros conjuntos de dados fossem gerados a partir de uma distribuição sem alteração. A partir do ponto de mudança, os valores passaram a ser $\delta = 0.1$ e $\gamma = 0.1$, respectivamente. Esses valores foram escolhidos com base em estudos prévios de (MAIA POLO *et al.*, 2023), que indicam que alterações dessa magnitude já são suficientes para induzir divergências detectáveis nas distribuições marginais e condicionais, sem tornar o problema trivial. Essa configuração permite avaliar de forma controlada a sensibilidade das medidas α_i^1 e α_i^2 , tanto em versões determinísticas quanto randomizadas dos *transducers* frente aos diferentes tipos de *shifts*.

O objetivo principal deste experimento é comparar a capacidade de detecção das duas medidas de não conformidade, analisando como cada uma delas impacta na sensibilidade do

método conformal frente às diferentes formas explícitas de *dataset shift*. As conclusões obtidas permitem verificar qual das medidas propostas apresenta maior robustez e precisão na detecção dessas mudanças distribucionais, contribuindo assim para uma melhor compreensão prática do desempenho dos métodos *conformal* baseados na divergência KL.

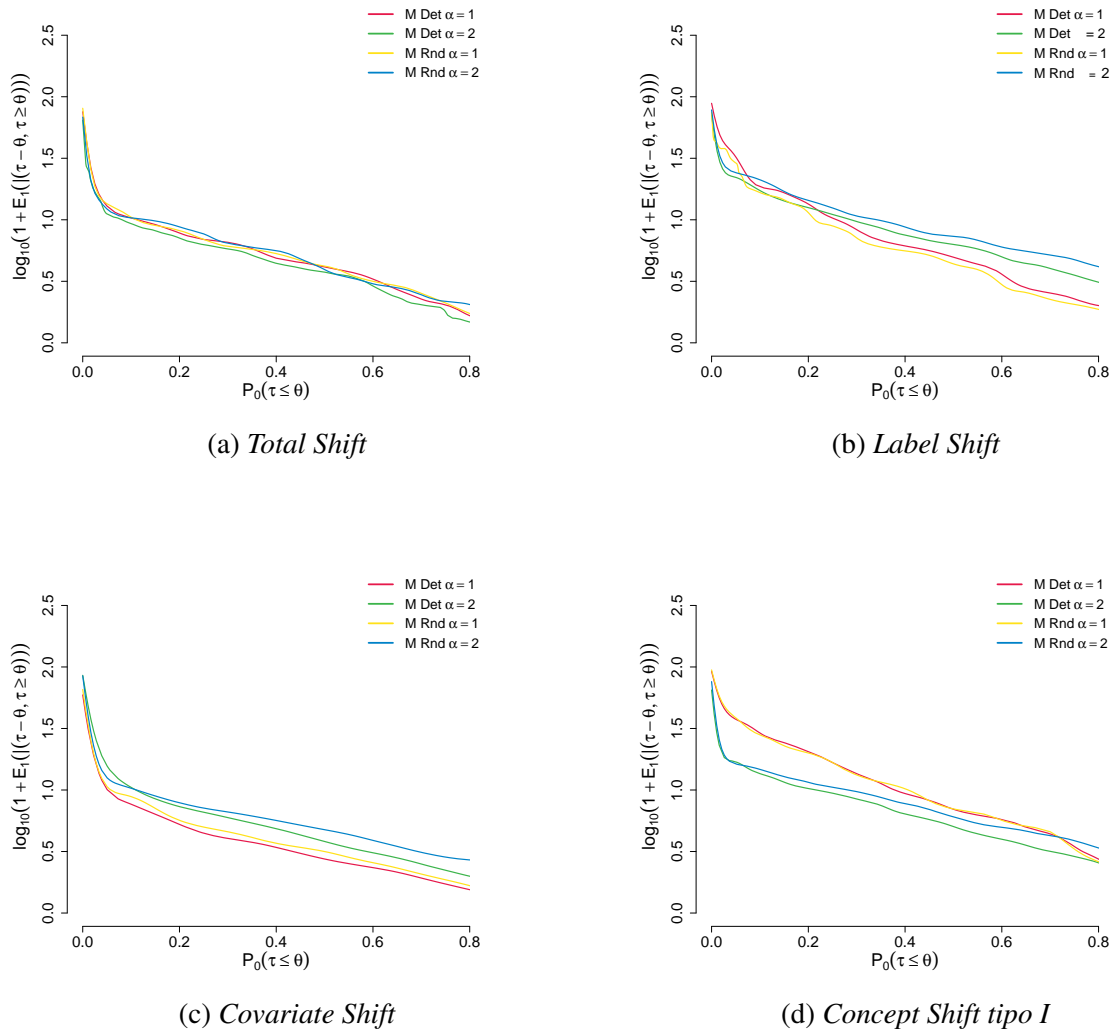


Figura 25 – Comparação entre as medidas de não conformidade e *transducers* para diferentes tipos de *shift* no Experimento.

Observando as Figuras 25 e 26, temos que para o *total shift*, observa-se um comportamento praticamente idêntico entre todas as combinações avaliadas de medida de não conformidade (α^1 e α^2) e *conformal transducer* (determinístico e randomizado). As curvas apresentam decaimento suave e quase sobreposto, com AUCs variando entre 0.39 e 0.41. Esse padrão indica que a divergência KL aplicada ao *total shift* consegue capturar a mudança e também que é sensível ao valor de limite que colocamos para detectar a mudança, independentemente da forma de agregação (*mínimo* ou *média*) e do tipo de *transducer* utilizado.

Nos cenários de *label shift*, *concept shift* tipo 1 e *covariate shift*, a discriminação entre as abordagens se torna um pouco mais evidente. Nos casos do *label shift* e do *covariate shift*, por

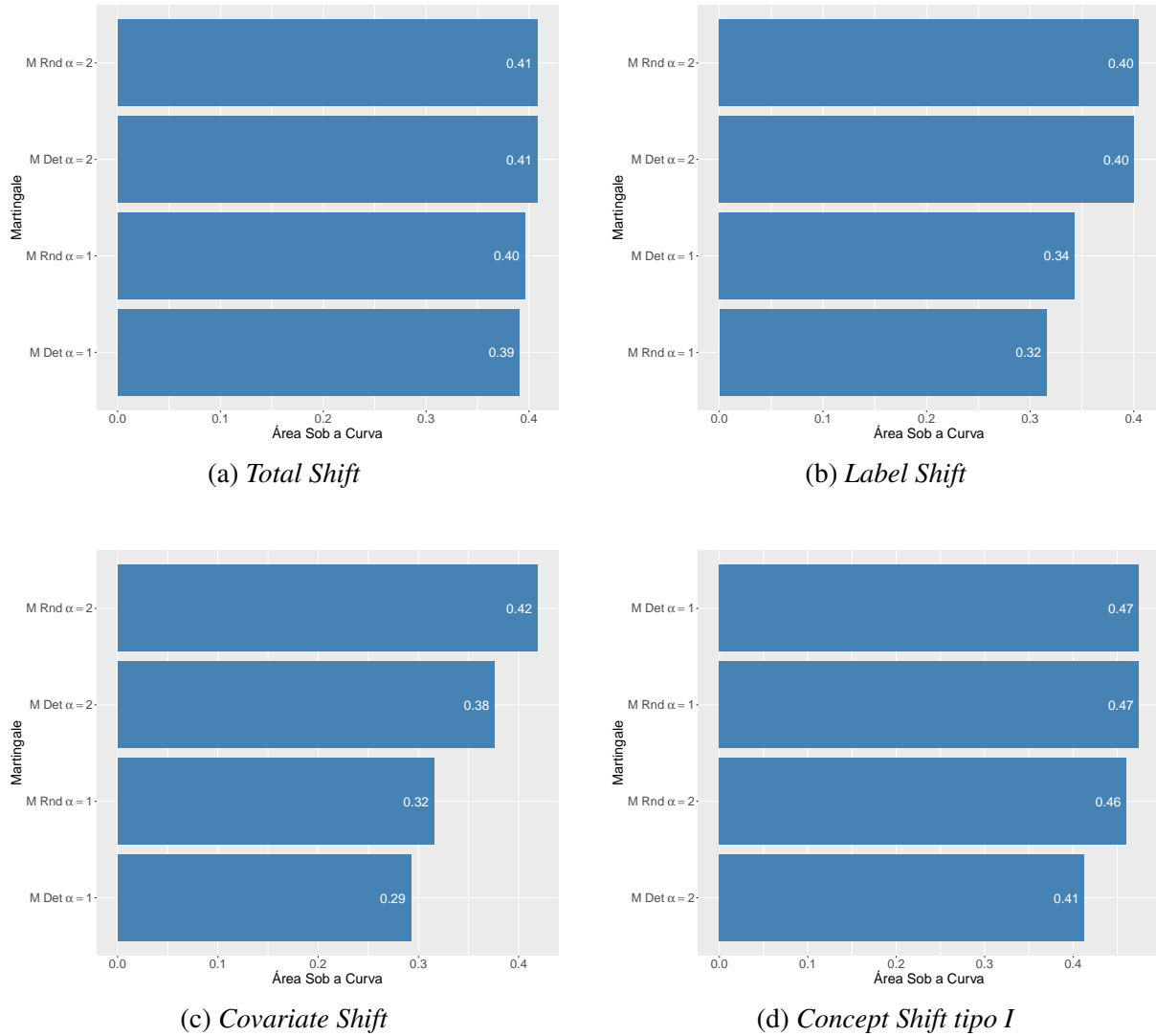


Figura 26 – AUCs obtidas para diferentes tipos de *shift* no Experimento.

exemplo, o uso de α^1 (mínimo das distâncias KL) mostrou-se mais eficaz do que α^2 (média), como ilustrado pelas AUCs menores do que quando utilizado α^2 . Já no *concept shift tipo I*, a diferença entre os métodos foi mais sutil, com destaque para os *transducers* determinístico e randomizado com α^2 atingindo as menores AUC's, enquanto α^1 teve leve queda de desempenho. Isso mostra que, nesse tipo de mudança, a escolha da medida de não conformidade impacta mais o desempenho do que o tipo de *transducer* em si, o oposto do observado no caso univariado e multivariado, em que o *transducer* randomizado frequentemente oferecia ganhos claros na maioria dos cenários.

Comparando com os experimentos anteriores (Capítulos 3 e 4), nota-se uma inversão de tendência em alguns casos. Por exemplo, anteriormente o *transducer* randomizado com α^2 frequentemente liderava o desempenho; no entanto, no *concept shift tipo I*, essa configuração ficou atrás do determinístico com α^1 . Observa-se que, no caso de *Total Shift*, embora as curvas resultantes das diferentes combinações de medidas de não conformidade e *transducers* apresentem

comportamentos muito similares, todas elas exibem um decaimento consistente à medida que a taxa de alarme falso aumenta. Dessa forma, mesmo na ausência de diferenças marcantes entre os métodos comparados e de métodos baseados em diferentes abordagens (como tínhamos nos casos univariados e multivariados), o próprio formato das curvas evidencia que o procedimento proposto — baseado na divergência de Kullback-Leibler combinada com *conformal* martingales — é eficaz para detectar mudanças globais na distribuição dos dados. Esse resultado reforça a viabilidade da abordagem, sobretudo em cenários nos quais se busca uma ferramenta geral e robusta para identificar desvios distribucionais ao longo do tempo.

Portanto, os resultados obtidos reforçam a aplicabilidade da medida de divergência KL como ferramenta de base para *conformal* martingales em cenários de *dataset shift*, apresentando desempenho consistente e robusto mesmo quando há pouca separabilidade entre os métodos.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Este trabalho explorou a detecção sequencial de mudanças por meio dos *Conformal Martingales*, passando das sequências univariadas mais simples até conjuntos de dados multivariados e lotes inteiros de observações. A investigação combinou teoria, simulações e comparações com métodos clássicos, sempre procurando responder a uma mesma pergunta: quão bem o método conformal se adapta quando o formato da mudança ou a própria estrutura dos dados variam?

Primeiro confirmamos, com dados sintéticos univariados, que as propriedades básicas previstas pela teoria aparecem na prática. Sob a hipótese nula os valores- p (que algumas vezes chamamos de p^c) exibem o comportamento uniforme (ou super-uniforme) exigido para que a taxa de alarmes falsos permaneça controlada; depois da mudança o martingale cresce de modo quase exponencial, produzindo um sinal claro de não conformidade. Esse resultado dá sustentação empírica à parte formal da literatura e justifica o uso do método nas etapas seguintes.

Em seguida confrontamos os *Conformal Martingales* com três testes clássicos de detecção de pontos de mudança utilizados na literatura — CUSUM, Shiryaev–Roberts e *Posterior Probability* — em vários cenários gerados a partir das distribuições Normal, de Poisson e Exponencial. Quando as distribuições antes e depois da mudança eram fornecidas de forma correta, os métodos clássicos mantiveram vantagem em atraso de detecção. No entanto bastou indicar um valor incorreto para a distribuição pós-mudança para que essa vantagem desaparecesse; em alguns casos o desempenho dos testes clássicos ficou claramente abaixo do obtido pelo martingale. Esse contraste reforça a mensagem central do trabalho: o método conformal sacrifica um pouco de velocidade, mas devolve robustez quando as premissas paramétricas deixam de ser confiáveis.

O estudo multivariado introduziu medidas de não conformidade que não haviam aparecido nos artigos de referência. Usamos profundidades estatísticas (L1, L2, potencial) e a distância de energia para avaliar mudanças na distribuição normal de dimensão dez. A profundidade L2 e a distância de energia mostraram ganhos pontuais em relação à distância euclidiana, sobretudo

quando a mudança ocorria na matriz de covariância em alta dimensão, enquanto a profundidade por potencial se revelou a alternativa mais conservadora. Esses resultados indicam que não existe uma única escolha universal: a medida de não conformidade deve levar em conta a geometria dos dados, o tipo de perturbação esperado e também alguma característica de interesse que justifique seu uso.

Por fim adaptamos o esquema conformal à detecção de *dataset shift* entre lotes sequenciais. Substituímos a distância entre pares de observações pela divergência de Kullback–Leibler entre conjuntos inteiros, estimada com o chamado *odds trick*. A mesma filosofia usada nos capítulos anteriores — utilizando *transducers* que geram valores-p, martingales determinístico e randomizado, e limites de perturbação variados — continuou válida, agora para mudanças de *total shift*, *label shift*, *covariate shift* e *concept shift* do tipo I. Embora não houvesse um teste clássico plenamente comparável, as curvas de tempo médio de detecção contra taxa de alarmes falsos mostraram padrões consistentes com aqueles observados nos experimentos univariados e multivariados, sinal de que a metodologia pode ser transferida para contextos de produção em que se monitoram fluxos de dados agregados.

Em resumo, a dissertação entregou quatro contribuições principais. A primeira foi a confirmação prática das garantias teóricas do método conformal. A segunda mostrou que a confiabilidade dos testes clássicos depende fortemente do conhecimento exato das distribuições envolvidas, enquanto os martingales mantêm estabilidade quando tal conhecimento falta ou é impreciso. A terceira ampliou o repertório de medidas de não conformidade, permitindo que o mesmo arcabouço trate dados de alta dimensão e distribuições com covariâncias complexas. A quarta adaptou a ideia para lotes de dados, utilizando a divergência KL para lidar com várias formas de *dataset shift* sem ajustes paramétricos adicionais.

Ainda há espaço para avanços. Por exemplo testes em fluxos reais, como em sensores industriais ou qualquer problema real que tenha natureza sequencial. Uma otimização dos códigos através de paralelização poderiam reduzir custos computacionais e permitir testar alguns cenários (principalmente os que utilizam a medida KL) de forma mais robusta e com mais variações.

Apesar dessas frentes em aberto, o trabalho deixa uma conclusão clara: os *Conformal Martingales* formam uma alternativa simples, flexível e útil para detecção online de mudanças, sobretudo quando o analista não dispõe de informações confiáveis sobre a distribuição após a mudança ou precisa monitorar diferentes tipos de desvio em um mesmo sistema. O método pode ser combinado com esquemas de retreinamento de modelos preditivos que se estende à indústria, à academia, a empresas de tecnologia, ao setor financeiro, etc.

REFERÊNCIAS

- ALIPPI, C.; ROVERI, M. Just-in-time adaptive classifiers—part i: Detecting nonstationary changes. **IEEE Transactions on Neural Networks**, v. 19, n. 7, p. 1145–1153, 2008. Citado na página 72.
- ATHREYA, K. B.; LAHIRI, S. N. **Measure Theory and Probability Theory**. New York: Springer, 2006. Citado na página 30.
- BENJAMINI, Y.; YEKUTIELI, D. The control of the false discovery rate in multiple testing under dependency. **Annals of Statistics**, Institute of Mathematical Statistics, v. 29, n. 4, p. 1165–1188, 2001. Citado na página 30.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, ASA Website, v. 65, n. 332, p. 1509–1526, 1970. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01621459.1970.10481180>>. Citado na página 42.
- COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2nd. ed. [S.l.]: John Wiley & Sons, 2006. Citado na página 73.
- CRANMER, K.; PAVEZ, J.; LOUPPE, G. **Approximating Likelihood Ratios with Calibrated Discriminative Classifiers**. 2016. Disponível em: <<https://arxiv.org/abs/1506.02169>>. Citado na página 74.
- DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression. iii. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 58, n. 1, p. 1–19, 1971. ISSN 00063444, 14643510. Disponível em: <<http://www.jstor.org/stable/2334313>>. Citado na página 42.
- GAMA, J.; ŽLIOBAITĚ, I.; BIFET, A.; PECHENIZKIY, M.; BOUCHACHIA, A. A survey on concept drift adaptation. **ACM Computing Surveys**, v. 46, n. 4, p. 44:1–44:37, 2014. Citado na página 71.
- KOSIOROWSKI, G.; MOSLER, K. **ddalpha: Depth- and Distance-Based Classification for Multivariate, Functional and Graph Data**. [S.l.], 2018. R package version 1.3.8. Disponível em: <<https://CRAN.R-project.org/package=ddalpha>>. Citado na página 65.
- KULLBACK, S.; LEIBLER, R. A. On Information and Sufficiency. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 22, n. 1, p. 79 – 86, 1951. Disponível em: <<https://doi.org/10.1214/aoms/1177729694>>. Citado na página 24.
- LIPTON, Z. C.; WANG, Y.-X.; SMOLA, A. Detecting and correcting for label shift with black box predictors. In: **Proceedings of the 35th International Conference on Machine Learning (ICML)**. [S.l.: s.n.], 2018. p. 3122–3130. Citado na página 71.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 08 1978. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/65.2.297>>. Citado na página 42.

- MAIA POLO, F.; IZBICKI, R.; LACERDA, E. G.; IBIETA-JIMENEZ, J. P.; VICENTE, R. A unified framework for dataset shift diagnostics. **Information Sciences**, v. 649, 2023. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025523011970>>. Citado na página 76.
- MORENO-TORRES, J. C.; RAEDER, T.; ALAIZ-RODRÍGUEZ, R.; CHAWLA, N. V.; HERRERA, F. A unifying view on dataset shift in classification. **Pattern Recognition**, v. 45, n. 1, p. 521–530, 2012. Citado na página 71.
- MORENO-TORRES, J. G.; RAEDER, T.; ALAIZ-RODRÍGUEZ, R.; CHAWLA, N. V.; HERRERA, F. A unifying view on dataset shift in classification. **Pattern Recognition**, v. 45, n. 1, p. 521–530, 2012. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320311002901>>. Citado na página 24.
- MOSLER, K. Depth statistics. In: BECKER, C.; GATHER, U.; FRIED, R. (Ed.). **Robustness and Complex Data Structures**. [S.l.]: Springer, 2013. p. 17–34. Citado nas páginas 24 e 65.
- MOUSTAKIDES, G. V. Optimal stopping times for detecting changes in distributions. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 14, n. 4, p. 1379–1387, 1986. Citado na página 51.
- PAGE, E. S. Continuous inspection schemes. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 41, n. 1/2, p. 100–115, 1954. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2333009>>. Citado nas páginas 24, 49 e 50.
- POLLAK, M. Optimal detection of a change in distribution. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 13, n. 1, p. 206–227, 1985. Citado na página 51.
- QUIÑONERO-CANDELA, J.; SUGIYAMA, M.; SCHWAIGHOFER, A.; LAWRENCE, N. D. (Ed.). **Dataset Shift in Machine Learning**. Cambridge, MA: MIT Press, 2009. (MIT Press). ISBN 9780262170055. Citado na página 71.
- RIZZO, M. L.; SZÉKELY, G. J. **energy: E-Statistics—Multivariate Inference via the Energy of Data**. [S.l.], 2023. R package version 1.7-13. Disponível em: <<https://CRAN.R-project.org/package=energy>>. Citado na página 65.
- ROBERTS, S. W. A comparison of some control chart procedures. **Technometrics**, ASA Website, v. 8, n. 3, p. 411–430, 1966. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00401706.1966.10490374>>. Citado na página 51.
- SCHERVISH, M. J. **Theory of Statistics**. New York, NY: Springer, 1995. (Springer Series in Statistics). ISBN 978-0387945460. Citado na página 27.
- SHIMODAIRA, H. Improving predictive inference under covariate shift by weighting the log-likelihood function. **Journal of Statistical Planning and Inference**, v. 90, n. 2, p. 227–244, 2000. Citado na página 71.
- SHIRYAEV, A. N. The problem of quickest detection of a violation of stationary behavior. **Dokl. Akad. Nauk SSSR**, v. 138, n. 5, p. 1039–1042, 1961. Disponível em: <<http://mi.mathnet.ru/dan25128>>. Citado nas páginas 24, 49 e 51.
- _____. On optimum methods in quickest detection problems. **Theory of Probability & Its Applications**, v. 8, n. 1, p. 22–46, 1963. Disponível em: <<https://doi.org/10.1137/1108002>>. Citado nas páginas 24, 49 e 52.

SUGIYAMA, M.; SUZUKI, T.; KANAMORI, T. **Density Ratio Estimation in Machine Learning**. [S.l.]: Cambridge University Press, 2012. Citado na página 74.

SZÉKELY, G. J.; RIZZO, M. L. Testing for equal distributions in high dimension. In: . [s.n.], 2004. Disponível em: <<https://api.semanticscholar.org/CorpusID:47751>>. Citado na página 59.

_____. **Testing for Equal Distributions in High Dimension**. [S.l.], 2004. Available at <<https://www.stat-energy.org>>. Citado na página 65.

VAPNIK, V. N. **Statistical Learning Theory**. New York: Wiley-Interscience, 1998. 768 p. Citado na página 31.

VARDI, Y.; ZHANG, C.-H. The multivariate l_1 -median and associated data depth. **Proceedings of the National Academy of Sciences**, v. 97, n. 4, p. 1423–1426, 2000. Disponível em: <<https://www.pnas.org/doi/abs/10.1073/pnas.97.4.1423>>. Citado na página 65.

VILLE, J. **Étude critique de la notion de collectif**. Paris: Gauthier-Villars, 1939. (Monographies des Probabilités, Fascicule III). Citado na página 33.

VOLKHONSKIY, D.; NOURETDINOV, I.; GAMMERMAN, A.; VOVK, V.; BURNAEV, E. **Inductive Conformal Martingales for Change-Point Detection**. 2017. Disponível em: <<https://arxiv.org/abs/1706.03415>>. Citado na página 49.

VOVK, V. Testing for concept shift online. **CoRR**, abs/2012.14246, 2020. Disponível em: <<https://arxiv.org/abs/2012.14246>>. Citado na página 40.

VOVK, V.; NOURETDINOV, I.; GAMMERMAN, A. Testing exchangeability on-line. In: FAWCETT, T.; MISHRA, N. (Ed.). **Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA**. AAAI Press, 2003. p. 768–775. Disponível em: <<http://www.aaai.org/Library/ICML/2003/icml03-100.php>>. Citado nas páginas 24, 27, 29, 30 e 36.

WIDMER, G.; KUBAT, M. Learning in the presence of concept drift and hidden contexts. **Machine Learning**, v. 23, n. 1, p. 69–101, 1996. Citado na página 72.

WU, Y. P. Y. **Information theory: From coding to learning**. [s.n.], 2022. Disponível em: <<https://people.lids.mit.edu/yp/homepage/papers.html>>. Citado na página 74.

DEDUÇÕES DAS FORMAS RECURSIVAS DOS MÉTODOS

Neste apêndice, apresentamos detalhadamente as deduções matemáticas que levam às expressões recursivas utilizadas para a implementação dos métodos CUSUM, Shiryaev–Roberts e Probabilidade Posterior (PP).

A.1 Método PP (Probabilidade Posterior)

A estatística ϕ_n é definida por:

$$e^{\phi_n} = \frac{p}{(1-p)^n} \sum_{i=1}^n \prod_{j=1}^{i-1} \frac{f_1(z_j)}{f_0(z_j)}, \quad n \geq 1.$$

Expandindo para $n + 1$:

$$e^{\phi_{n+1}} = \frac{p}{(1-p)^{n+1}} \sum_{i=1}^{n+1} \prod_{j=1}^{i-1} \frac{f_1(z_j)}{f_0(z_j)}.$$

Separando o último termo ($i = n + 1$):

$$e^{\phi_{n+1}} = \frac{p}{(1-p)^{n+1}} \left(\sum_{i=1}^n \prod_{j=1}^{i-1} \frac{f_1(z_j)}{f_0(z_j)} + \prod_{j=1}^n \frac{f_1(z_j)}{f_0(z_j)} R(z_{n+1}) \right).$$

Colocando $R(z_j) = \frac{f_1(z_j)}{f_0(z_j)}$, temos:

$$e^{\phi_{n+1}} = \frac{p}{(1-p)^{n+1}} \left(\sum_{i=1}^n \prod_{j=1}^{i-1} R(z_j) + \prod_{j=1}^n R(z_j) R(z_{n+1}) \right).$$

Distribuindo:

$$e^{\phi_{n+1}} = \frac{1}{1-p} \left(\frac{p}{(1-p)^n} \sum_{i=1}^n \prod_{j=1}^{i-1} R(z_j) \right) R(z_{n+1}) + \frac{p}{(1-p)^{n+1}} R(z_{n+1}).$$

Substituindo e^{ϕ_n} :

$$e^{\phi_{n+1}} = \frac{1}{1-p} \left(e^{\phi_n} R(z_{n+1}) + p R(z_{n+1}) \right),$$

$$e^{\phi_{n+1}} = \frac{R(z_{n+1})}{1-p} (e^{\phi_n} + p).$$

Aplicando o logaritmo:

$$\phi_{n+1} = \log(R(z_{n+1})) - \log(1-p) + \log(p + e^{\phi_n}).$$

Além disso, o valor inicial é:

$$\phi_1 = \log(p) - \log(1-p) + \log(R(z_1)).$$

—

A.2 Método CUSUM

A estatística γ_n do método CUSUM é dada por:

$$\gamma_n = \max_{1 \leq i \leq n} \sum_{j=i}^n \log(R(z_j)).$$

Sabemos que:

$$\gamma_1 = \log(R(z_1)).$$

Agora, para $n+1$:

$$\gamma_{n+1} = \max_{1 \leq i \leq n+1} \sum_{j=i}^{n+1} \log(R(z_j)).$$

Observamos que:

$$\sum_{j=i}^{n+1} \log(R(z_j)) = \sum_{j=i}^n \log(R(z_j)) + \log(R(z_{n+1})).$$

Assim:

$$\gamma_{n+1} = \max \left(\max_{1 \leq i \leq n} \left(\sum_{j=i}^n \log(R(z_j)) + \log(R(z_{n+1})) \right), \log(R(z_{n+1})) \right).$$

Organizando:

$$\gamma_{n+1} = \log(R(z_{n+1})) + \max(\gamma_n, 0).$$

A.3 Método Shiryayev–Roberts

A estatística ψ_n do método Shiryayev–Roberts é definida por:

$$\psi_n = \log \left(\sum_{i=1}^n \prod_{j=i}^n R(z_j) \right).$$

Expandindo:

$$e^{\psi_n} = \sum_{i=1}^n \prod_{j=i}^n R(z_j).$$

No passo $n + 1$:

$$e^{\psi_{n+1}} = \sum_{i=1}^n \left(\prod_{j=i}^n R(z_j) \right) R(z_{n+1}) + R(z_{n+1}),$$

$$e^{\psi_{n+1}} = R(z_{n+1}) \left(\sum_{i=1}^n \prod_{j=i}^n R(z_j) + 1 \right),$$

$$e^{\psi_{n+1}} = R(z_{n+1}) (e^{\psi_n} + 1).$$

Aplicando o logaritmo:

$$\psi_{n+1} = \log(R(z_{n+1})) + \log(e^{\psi_n} + 1),$$

com condição inicial:

$$\psi_1 = \log(R(z_1)).$$

