

Teste de hipóteses com estrutura combinatorial: problemas de detecção de médias e correlações

Helder Jorge Quitério Santiago

Orientador: Guilherme Ost



Universidade Federal do Rio de Janeiro
Instituto de Matemática
Departamento de Métodos Estatísticos

Rio de Janeiro 2020

Teste de hipóteses com estrutura combinatorial: problemas de detecção de médias e correlações

Helder Jorge Quitério Santiago

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

TESTE DE HIPÓTESES COM ESTRUTURA
COMBINATORIAL: PROBLEMAS DE DETECÇÃO DE
MÉDIAS E CORRELAÇÕES

Helder Jorge Quitério Santiago

Orientador: Guilherme Ost

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Guilherme Ost
IM-UFRJ

Giulio Iacobelli
IM-UFRJ

Florencia Leonardi
IME-USP

Rio de Janeiro 2020

Dedico esse trabalho à meus pais, minha irmã e meus avós que sempre estiveram ao meu lado em todos os momentos da minha caminhada dando apoio e carinho que foram essenciais para que eu chegasse até aqui.

Agradecimentos

Agradeço à minha mãe, meu pai, minha irmã e meus avós pelos ensinamentos, encorajamentos e apoio durante toda a minha vida, e em especial durante o meu mestrado. Ao meu orientador, professor Guilherme Ost de Aguiar, pela paciência, apoio, dedicação, generosidade e amizade que contribuíram fortemente na construção deste trabalho. Aos professores do DME-IM - Departamento de Matemática e Estatística do Instituto de Matemática, pelos ensinamentos dados por eles nas disciplinas cursadas ao longo da pós-graduação e à toda a equipe do DME-IM e a Capes que proporcionaram as condições necessárias para a realização do meu mestrado. Por fim, agradeço à Deus que me deu força e sabedoria até a finalização deste estudo.

Saber Viver

*“Não sei...
se a vida é curta
ou longa demais para nós.
Mas sei que nada do que vivemos
tem sentido,
se não tocarmos o coração das pessoas.*

*Muitas vezes basta ser:
colo que acolhe,
braço que envolve,
palavra que conforta,
silêncio que respeita,
alegria que contagia,
lágrima que corre,
olhar que sacia,
amor que promove.*

*E isso não é coisa de outro mundo:
é o que dá sentido à vida.*

*É o que faz com que ela
não seja nem curta,
nem longa demais,
mas que seja intensa,
verdadeira e pura...
enquanto durar.”*

Cora Coralina

Resumo

Nesta dissertação, estudamos o problema de detectar anomalias do ponto de vista de testes de hipóteses. Mais especificamente, estudamos dois problemas de detecção no contexto Gaussiano, considerados em (Addario-Berry et al. (2010)) e (Arias-Castro et al. (2012)), onde as variáveis com anomalias têm alguma estrutura geométrica. O primeiro problema é chamado detecção de médias. Nesse problema, observamos um vetor de alta dimensão com coordenadas independentes e queremos decidir se todas as coordenadas possuem distribuição normal padrão ou, alternativamente, se existe um subconjunto de coordenadas (pertencendo a uma dada classe de conjuntos) para as quais a distribuição é uma normal com média diferente de zero e variância um. Nessa formulação, cada coordenada do vetor aleatório cuja média da distribuição normal correspondente é não nula é chamada anômala.

O segundo problema é chamado detecção de correlações. Nesse problema, queremos decidir se um vetor de alta dimensão possui coordenadas normais padrão independentes ou, alternativamente, se existe um subconjunto de coordenadas (pertencendo a uma dada classe de conjuntos) que são correlacionadas. Nessa formulação, as coordenadas do vetor aleatório correlacionadas são chamadas anômalas.

Em ambos os problemas, buscamos entender quando é possível e impossível detectar a presença de variáveis anômalas.

Palavras chave: Teste de Hipóteses, Detecção de médias, Detecção de Correlações, Risco de Bayes, Risco Minimax

Abstract

In this dissertation, we study anomaly detection problems from the point of view of hypothesis testing. More specifically, we study two detection problems in the Gaussian context, considered in ([Addario-Berry et al. \(2010\)](#)) and ([Arias-Castro et al. \(2012\)](#)), where the variables with anomalies have some geometric structure. The first problem is called detection of means. In this problem, we observe a high-dimensional vector with independent coordinates and want to decide whether all coordinates follow standard normal distribution or, alternatively, if there is a subset of coordinates (belonging to a given class of sets) for which the distribution is a normal with a non-zero mean and variance one. In this formulation, each coordinate of the random vector whose mean of the corresponding normal distribution is non-zero is called anomalous.

The second problem is called detection of correlations. In this problem, we want to decide whether a high-dimensional vector has independent standard normal coordinates or, alternatively, whether there is a subset of coordinates (belonging to a given class of sets) that are correlated. In this formulation, the coordinates of the correlated random vector are called anomalous.

In both problems, we seek to understand when it is possible and impossible to detect the presence of anomalous variables.

Keywords: Hypothesis Testing, Detection of Means, Detection of Correlations, Bayes Risk, Minimax Risk

Conteúdo

1	Introdução	13
2	Formulação dos problemas de detecção	19
3	Detecção de médias	29
3.1	Cotas superiores	29
3.1.1	Teste baseado na soma	29
3.1.2	Teste baseado no máximo	31
3.2	Cotas Inferiores	35
3.2.1	Cota inferior universal	35
3.2.2	Cota inferior baseada em pares sobrepostos	37
3.2.3	Cotas inferiores para classes simétricas	39
3.3	Exemplos	43
3.3.1	Conjuntos disjuntos	43
3.3.2	k_n -conjuntos	45
3.3.3	Estrelas	49
3.3.4	Árvores geradoras	52
3.4	Monotonicidade do risco de Bayes	56
3.5	Cota inferior baseada em subclasses aleatórias	59
3.6	Uma análise parcial do teste ótimo contra o teste baseado no máximo	68
4	Detecção de correlações	74
4.1	Cotas inferiores	75
4.1.1	Caso $N_n = 1$	75
4.1.2	Cota inferior baseada em pares sobrepostos	81
4.2	Exemplos	84
4.2.1	Conjuntos disjuntos	85
4.2.2	k_n -conjuntos	87
4.2.3	k_n -intervalos	89
4.2.4	Árvores geradoras	91

4.3	Alguns testes auxiliares	93
4.3.1	Teste do Quadrado da Soma	93
4.3.2	Teste da Razão de Verossimilhança Generalizado	95
4.3.3	Teste do Quadrado da Soma localizada	105
4.3.4	Teste da Bondade de Ajuste	107
A	Lema de Neyman-Pearson	114
B	Lema de Slepian	117
C	Lema de Tsirelson	122
D	Teorema de Bernstein	125

Capítulo 1

Introdução

Frequentemente, ao se analisar conjuntos de dados do mundo real é fundamental determinar quais elementos desse conjunto apresentam um comportamento *anômalo*, isto é, diferente do *esperado*. Na literatura, esse problema é conhecido como *detecção de anomalias* ou ainda *detecção de outliers* (Chalapathy and Chawla (2019)).

Nessa dissertação abordaremos o problema de detecção de anomalias sob o ponto de vista de *teste de hipóteses*. Nesse contexto, uma formulação para o problema de detecção frequentemente adotada é a seguinte: observamos $n \geq 1$ variáveis aleatórias independentes e queremos testar se essas variáveis têm uma mesma função de distribuição acumulada f.d.a F_0 ou, alternativamente, se uma pequena parcela ϵ_n delas têm uma f.d.a diferente $F_1^{(n)}$. Nessa formulação, as variáveis aleatórias com f.d.a $F_1^{(n)}$ apresentariam um comportamento anômalo. Note que $F_1^{(n)}$ pode variar com o número n de variáveis observadas. Na literatura, esse problema é chamado de *detecção de misturas esparsas* (Arias-Castro and Ying (2019)) e possui aplicações em diversas áreas do conhecimento, incluindo processamento de sinais (Dobrusin (1958)), bioestatística (Jeng et al. (2010)) e astrofísica (Cayon et al. (2005) e Jin et al. (2005)).

Em (Ingster (1996)), o problema de detecção de misturas esparsas foi tratado no caso *Gaussiano*, onde F_0 corresponde a f.d.a de uma distribuição normal padrão e $F_1^{(n)}$ corresponde a f.d.a de uma distribuição normal com média μ_n e variância 1. Nesse contexto, o problema de detecção de misturas esparsas pode ser reformulado do seguinte modo: quando é possível testar a hipótese nula de que as médias de n normais de variância 1 são todas zero contra a alternativa de que uma pequena fração ϵ_n tem média μ_n não nula?

Pode-se verificar que quando os parâmetros ϵ_n e μ_n são conhecidos, o teste da razão de verossimilhança é o teste mais poderoso. Esse fato é uma consequência do Lema de Neyman-Pearson já que as hipóteses nula e alter-

nativa são do tipo simples quando ϵ_n e μ_n são conhecidos. Os resultados de (Ingster (1996)) nos dizem que ao tomar $\epsilon_n = n^{-\beta}$ e $\mu_n = \sqrt{2r \log n}$ onde $\frac{1}{2} < \beta < 1$ e $0 < r < 1$, e definindo a função

$$\rho^*(\beta) = \begin{cases} \beta - \frac{1}{2}, & \text{se } \frac{1}{2} < \beta \leq \frac{3}{4}, \\ (1 - \sqrt{1 - \beta})^2, & \text{se } \frac{3}{4} < \beta < 1, \end{cases}$$

o seguinte fenômeno (chamado *efeito limiar*) ocorre:

- Se $r > \rho^*(\beta)$, então a soma das probabilidades dos erros do tipo 1 e tipo 2 para o teste da razão de verossimilhança converge para 0 quando o tamanho da amostra $n \rightarrow \infty$.
- Se $r < \rho^*(\beta)$, então a soma das probabilidades dos erros do tipo 1 e tipo 2 para o teste da razão de verossimilhança converge para 1 quando o tamanho da amostra $n \rightarrow \infty$.

O resultado acima pode ser interpretado da seguinte maneira. A curva dada por $(1/2, 1) \ni \beta \mapsto (\beta, \rho^*(\beta))$ particiona o plano (β, r) com $1/2 < \beta < 1$ e $0 < r < 1$ em duas regiões complementares. Na região definida por (β, r) com $r > \rho^*(\beta)$ é possível detectar a presença de variáveis anômalas com o teste da razão de verossimilhança, enquanto que na região (β, r) com $r < \rho^*(\beta)$ é impossível detectar a presença de variáveis anômalas, seja qual for o teste estatístico escolhido. Tais efeitos limiares foram subsequentemente investigados para outros modelos. Por exemplo, (Cai and Wu (2014)) trataram do problema de detecção de misturas esparsas Gaussianas em que a variância das variáveis anômalas podem ser diferentes de 1 e extensões para modelos não Gaussianos foram estudados em (Donoho and Jin (2004)).

Uma segunda formulação para o problema de detecção de anomalias também estudada é a seguinte. Observamos um vetor aleatório n -dimensional X cujas coordenadas são independentes. A partir dessa observação, queremos decidir se todas as coordenadas do vetor aleatório têm distribuição normal padrão ou, alternativamente, se existe um subconjunto S de $k_n \geq 1$ coordenadas (pertencendo a uma dada classe de conjuntos \mathcal{C}_n) para as quais a distribuição é uma normal de média $\mu_n > 0$ e variância 1. Nessa formulação, dizemos que uma coordenada do vetor X apresenta um comportamento anômalo se a média da distribuição normal correspondente é diferente de zero, e nesse caso é necessariamente dada por μ_n . Na literatura, esse problema é chamado de *detecção de médias*.

Uma diferença significativa entre os problemas de detecção de médias e de misturas esparsas é que no primeiro há uma *estrutura combinatorial* entre as variáveis anômalas, representada pela classe de conjuntos \mathcal{C}_n , que não está

presente no segundo. Tipicamente, as coordenadas do vetor X representam pesos em dado grafo G de n arestas e \mathcal{C}_n é uma classe de subgrafos de G .

O problema de detecção de médias foi introduzido por (Arias-Castro et al. (2008)), onde dois exemplos para a classe \mathcal{C}_n foram estudados em detalhe. No primeiro exemplo, \mathcal{C}_n contém todos os caminhos de tamanho n entre dois dados vértices em um grid bidimensional e no segundo exemplo, \mathcal{C}_n é o conjunto de todos os caminhos de uma raiz até uma folha em uma árvore binária completa de altura n . Subsequentemente em (Arias-Castro et al. (2009)), foi investigado a classe de exemplos onde elementos de \mathcal{C}_n correspondem a clusters em um grid regular. Outros exemplos para a classe \mathcal{C}_n , incluindo *árvores geradoras* e *estrelas* de um grafo completo, foram considerados em (Addario-Berry et al. (2010)). Possíveis aplicações para o problema de detecção de médias podem ser encontradas em vigilância baseada em redes de sensores (Culler et al. (2004)) e detecção de materiais substâncias radioativas (Brennan et al. (2004)), além da detecção de sinais (Arias-Castro et al. (2008)).

Uma terceira formulação para o problema de detecção de anomalias, chamada de *problema de detecção de correlações* e introduzida por (Arias-Castro et al. (2012)), pode ser descrita da seguinte maneira. Observamos um vetor aleatório n -dimensional X com distribuição *normal multivariada* com *vetor de médias* $\mathbf{0}_n$ (vetor n -dimensional com todas as entradas iguais nulas) e *matriz de covariância* Σ_n . Com base nessa observação queremos decidir se todas as coordenadas do vetor aleatório são independentes (i.e., Σ_n é a matriz identidade \mathbf{I}_n de tamanho n) ou, alternativamente, se existe um subconjunto S de $k_n \geq 1$ coordenadas (pertencendo a uma dada classe de conjuntos \mathcal{C}_n) para as quais a correlação entre elas é $\rho_n > 0$, isto é, as entradas da matriz de covariância são $(\Sigma_n)_{i,j} = \rho_n$ para todo $i \neq j$ com $i, j \in S$, $(\Sigma_n)_{i,j} = 1$ para todo $i = j$ e $(\Sigma_n)_{i,j} = 0$ para $i \neq j$ com $i, j \in S^c$. Nessa formulação, dizemos que as coordenadas do vetor X correlacionadas apresentam um comportamento anômalo.

Há uma diferença crucial entre o problema de detecção de correlações e os outros dois problemas anteriores. De fato, no problema de detecção de correlações uma parcela das variáveis (i.e., das coordenadas do vetor aleatório X) são *dependentes* sob a hipótese alternativa, enquanto que sob a mesma hipótese nos outros dois problemas as variáveis são sempre *independentes*. Essa diferença torna o estudo do problema de detecção de correlações mais delicado. Possíveis aplicações para o problema de detecção de correlações, bem como a sua conexão com outros problemas em estatística podem ser encontradas em (Arias-Castro et al. (2012)).

Objetivos da dissertação. Essa dissertação tem como objetivo estudar os limites de detecção para os problemas de detecção de médias e correlações.

Em outras palavras, buscamos entender, de modo similar ao feito em (Ings-ter (1998)), quando é possível e impossível detectar a presença de variáveis anômalas para cada um desses problemas de detecção. Para esse fim, seguimos os seguintes passos. No primeiro passo, determinamos quais são os testes estatísticos *ótimos*. De maneira informal, o risco de um teste é uma medida de desempenho do teste: quanto menor o risco, melhor é o teste. Dessa maneira, dizemos que um teste estatístico é ótimo se o seu *risco* é o menor entre todos os testes possíveis. No segundo passo, buscamos estabelecer condições sobre os parâmetros (o tamanho da amostra n , a estrutura da classe \mathcal{C}_n considerada, o tamanho k_n de cada subconjunto S em \mathcal{C}_n e o valor μ_n no problema de detecção de médias - o valor ρ_n no problema de detecção de correlações) de modo que o risco do teste ótimo seja pelo menos δ_n , para cada $\delta_n \in (0, 1)$ fixado. Esses resultados são chamados de cotas inferiores para o risco do teste ótimo. Terceiro, estabelecemos condições complementares sobre os mesmos parâmetros de modo que o risco do teste ótimo seja no máximo δ_n . Esses resultados são chamados de cotas superiores para o risco do teste ótimo. Obter cotas superiores diretamente para o risco do teste ótimo é em geral difícil e, por isso, consideramos testes estatísticos auxiliares. Cotas superiores para o risco desses testes implicam automaticamente cotas superiores para o risco do teste ótimo. Por fim, utilizamos as cotas inferiores (respectivamente superiores) para determinar quando é impossível (respectivamente possível) detectar a presença de variáveis anômalas. A análise feita nessa dissertação se concentra nos riscos *médio* e *do pior caso possível*. De maneira informal, o risco médio é a soma da probabilidade de se rejeitar a hipótese nula dado que a hipótese nula é verdadeira e da média aritmética das probabilidades de se aceitar a hipótese nula dado que existe um conjunto de coordenadas $S \in \mathcal{C}_n$ cuja média é diferente de zero. De maneira similar o risco do pior caso possível é a soma da probabilidade de se rejeitar a hipótese nula dado que a hipótese nula é verdadeira e do máximo sob todos os conjuntos $S \in \mathcal{C}_n$ da probabilidade de se aceitar a hipótese nula dado que existe um conjunto de coordenadas $S \in \mathcal{C}_n$ cuja média é diferente de zero. Embora diferentes, é importante destacar que em muitos dos exemplos considerados os riscos médio e do pior caso possível para o teste ótimo coincidem. Os resultados aqui apresentados são não assintóticos (isto é, são válidos para todo n fixado) e são obtidos via *desigualdades de concentração*. Apesar disso, destacamos que em cada exemplo estudado, investigamos o comportamento assintótico das cotas inferiores e superiores obtidas para o risco do teste ótimo. Em alguns casos, mostramos que as cotas inferiores e superiores têm a mesma ordem de magnitude, isto é, a menos de constante elas têm a mesma expressão. Nesses casos, temos um entendimento completo dos problemas de detecções. Enfatizamos que os problemas aqui estudados foram considerados

em (Addario-Berry et al. (2010)) e (Arias-Castro et al. (2012)).

Contribuição da dissertação. A primeira contribuição dessa dissertação é apresentar um texto introdutório e auto contido aos problemas de teste de hipóteses com estrutura combinatorial considerados em (Addario-Berry et al. (2010)) e (Arias-Castro et al. (2012)). Os resultados apresentados nesses artigos para os exemplos lá considerados são formulados sob pontos de vista distintos. Mais especificamente, no artigo (Addario-Berry et al. (2010)) esses resultados são formulados sob o ponto de vista não assintótico, enquanto que em (Arias-Castro et al. (2012)) os resultados correspondentes são formulados sob o ponto de vista assintótico. A segunda contribuição dessa dissertação é, portanto, formular esses resultados de ambos os artigos em um mesmo quadro conceitual, que é sob o ponto de vista assintótico.

Organização do restante do texto. No Capítulo 2, introduzimos a notação utilizada na dissertação, formulamos matematicamente os problemas de detecção de médias e de correlações, explicitamos exemplos da classe \mathcal{C}_n estudados e provamos alguns resultados preliminares. Um desses resultados é a Proposição 1 que afirma que o *teste da razão de verossimilhança* (TRV) é *ótimo* no sentido de que o seu risco médio é o menor entre todos os testes possíveis.

No Capítulo 3 nos concentramos no problema de detecção de médias. Na Seção 3.1, estudamos em detalhe dois testes estatísticos auxiliares (o *teste baseado no máximo* e o *teste baseado na soma*) que são úteis em nossa análise para determinar cotas superiores para o risco médio do TRV. Na Seção 3.2, obtemos cotas inferiores para o risco médio do TRV baseadas na função geradora de momentos da sobreposição de dois conjuntos escolhidos de maneira uniforme em \mathcal{C}_n e de forma independente. Na Seção 3.3, aplicamos os resultados das Seções 3.1 e 3.2 em algumas classes \mathcal{C}_n para determinar quando é possível e impossível detectar anomalias. Além disso, na Seção 3.4 mostramos que o risco médio não é, em geral, monótono no sentido de que classes \mathcal{C}_n maiores podem ser significativamente mais fáceis de testar do que classes menores, embora a monotonicidade se mantenha sob certas condições de simetria. Na Seção 3.5, apresentamos outra cota inferior para o risco médio do TRV baseada na mediana da maior sobreposição entre subconjuntos de \mathcal{C}_n escolhidos ao acaso e *sem reposição*. Finalizamos o Capítulo 3, com a Seção 3.6 na qual fazemos uma análise parcial do TRV contra o teste baseado no máximo.

No Capítulo 4, nos concentramos no problema de detecção de correlações. Na Seção 4.1, obtemos cotas inferiores para o risco médio do TRV em termos da função geradora de momentos do tamanho da sobreposição de dois elementos escolhidos aleatoriamente da classe \mathcal{C}_n . Na Seção 4.2, aplicamos as cotas inferiores em algumas classes \mathcal{C}_n para determinar, sob que condições

dos parâmetros, é impossível detectar anomalias. Na Seção 4.3, estudamos o desempenho de alguns testes estatísticos auxiliares como o *teste do quadrado da soma*, o *teste da razão da verossimilhança generalizado* e o *teste da bondade de ajuste*, determinando, para algumas das classes \mathcal{C}_n consideradas na Seção 4.2, condições nos parâmetros sob as quais é possível detectar anomalias. Além disso, indicamos regimes sob os quais o risco do pior caso possível para o teste ótimo e o risco do pior caso possível para o teste auxiliar têm o mesmo comportamento assintótico. Nesses casos, o teste auxiliar tem o mesmo desempenho do teste ótimo. No Apêndice demonstramos resultados importantes usados ao longo de toda a dissertação, a saber, os Lemas de Neyman-Pearson, Slepian e Tsirelson, assim como o Teorema de Bernstein.

Capítulo 2

Formulação dos problemas de detecção

Notação. Sejam $(a_m)_{m \geq 1}$ e $(b_m)_{m \geq 1}$ duas seqüências de números reais. Dizemos que a_m é assintoticamente igual b_m , e nesse caso escrevemos $a_m \approx b_m$, se $a_m/b_m \rightarrow 1$ quando $m \rightarrow \infty$. Além disso, escrevemos $a_m \gg b_m$ se $a_m/b_m \rightarrow \infty$, quando $m \rightarrow \infty$. Dados os conjuntos A_1, \dots, A_n e $I \subset \{1, \dots, m\}$, denotamos por $\prod_{i \in I} A_i$ o produto cartesiano dos conjuntos A_i cujo índice pertence ao conjunto I . Sejam Y, Y_1, Y_2, \dots variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$. Dizemos que Y_m converge para Y *quase certamente* se $\mathbb{P}\{Y_m \rightarrow Y \text{ quando } m \rightarrow \infty\} = 1$, isto é, se o evento $A_0 = \{\omega \in \Omega : Y_n(\omega) \rightarrow Y(\omega)\}$ é de probabilidade 1. Denotamos $Y_m \xrightarrow[m \rightarrow \infty]{q.c.} Y$ para indicar que Y_n converge quase certamente para Y quando $n \rightarrow \infty$. Sejam Z, Z_1, Z_2, \dots variáveis aleatórias com funções de distribuição F, F_1, F_2, \dots , respectivamente. Dizemos que Z_m converge em *distribuição* para Z , quando $m \rightarrow \infty$, se $F_m(z) \rightarrow F(z)$ para todo z ponto de continuidade de F . Denotamos $Z_m \xrightarrow{\mathcal{D}} Z$ para indicar que Z_m converge em distribuição para Z quando $m \rightarrow \infty$.

Seja $n \geq 1$ um inteiro fixado. Para todos os vetores $x, y \in \mathbb{R}^n$, denotamos por $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ o produto interno usual em \mathbb{R}^n e $\|x\|_2 = \sqrt{\langle x, x \rangle}$. Denotamos por $\mathbf{0}_n$ o vetor n -dimensional com todas as suas coordenadas iguais a zero e \mathbf{I}_n a matriz identidade $n \times n$. Além disso, durante toda a dissertação, escreveremos $Y \sim \mathcal{N}(\mathbf{v}, \Sigma)$ para indicar que o vetor aleatório $Y = (Y_1, \dots, Y_n)$ segue uma distribuição normal multivariada com vetor de médias $\mathbf{v} = (v_1, \dots, v_n)$ e matriz de covariâncias $\Sigma = (\Sigma_{i,j})_{1 \leq i, j \leq n}$. Para cada matriz $\Sigma = (\Sigma_{ij})_{1 \leq i, j \leq n}$ de entradas reais denotamos Σ^T como sendo a matriz transposta de Σ . Uma matriz B é dita ser raiz quadrada de A e denotada por $A^{1/2}$ se o produto BB é igual a A .

No que segue, consideramos um vetor aleatório $X = (X_1, \dots, X_n)$ e uma classe de subconjuntos $\mathcal{C}_n = \{S_1, \dots, S_{N_n}\}$ tal que $S_i \subset \{1, \dots, n\}$ com $|S_i| = k_n$, isto é, cada subconjunto S_i tem k_n elementos.

Detecção de médias. A partir da observação do vetor aleatório X , queremos testar a hipótese nula,

$$H_0 : X \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n),$$

contra a hipótese alternativa,

$$H_1 : X \sim \mathcal{N}(\mu_S, \mathbf{I}_n) \text{ para algum } S \in \mathcal{C}_n.$$

Na hipótese H_1 acima, $\mu_S = ((\mu_S)_1, \dots, (\mu_S)_n)$ é um vetor n -dimensional definido por

$$(\mu_S)_i = \begin{cases} 0, & \text{se } i \notin S, \\ \mu_n, & \text{se } i \in S, \end{cases} \quad (2.1)$$

onde $\mu_n > 0$ é um parâmetro conhecido e positivo.

Denotamos \mathbb{P}_0 a medida de probabilidade sob a qual $X \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$ e \mathbb{P}_S a medida de probabilidade sob a qual $X \sim \mathcal{N}(\mu_S, \mathbf{I}_n)$ onde μ_S é o vetor definido em (2.1). Além disso, denotamos \mathbb{E}_0 e \mathbb{E}_S os valores esperados em relação as medidas de probabilidade \mathbb{P}_0 e \mathbb{P}_S , respectivamente.

Detecção de correlações. Com base no vetor aleatório X , queremos testar a hipótese nula

$$H_0 : X \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n), \quad (2.2)$$

contra a hipótese alternativa

$$H_1 : X \sim \mathcal{N}(\mathbf{0}_n, A_S) \text{ para algum } S \in \mathcal{C}_n.$$

Na hipótese H_1 acima, A_S é uma matriz $n \times n$ cuja entrada $(A_S)_{i,j}$ é dada por

$$(A_S)_{i,j} = \begin{cases} 1, & \text{se } i = j, \\ \rho_n, & \text{se } i \neq j, \text{ com } i, j \in S, \\ 0, & \text{caso contrário,} \end{cases} \quad (2.3)$$

onde $\rho_n > 0$ é um parâmetro dado.

No problema de detecção de correlações, denotaremos \mathbb{P}_S a medida de probabilidade sob a qual $X \sim \mathcal{N}(\mathbf{0}_n, A_S)$ onde A_S é a matriz definida em (2.3).

Nessa dissertação estudaremos os seguintes exemplos para a classe \mathcal{C}_n :

- k_n -conjuntos. Nesse exemplo, consideramos a classe de todos os conjuntos de tamanho k_n , isto é, todo subconjunto de índices S é da forma $\{i_1, \dots, i_{k_n}\} \subset \{1, \dots, n\}$, onde i_1, \dots, i_{k_n} são todos distintos. Nesse caso, $N_n = \binom{n}{k_n}$.
- k_n -intervalos. Nesse exemplo, consideramos a classe de todos os intervalos de tamanho k_n . Suponha que $n = Lk_n$ para algum inteiro L positivo. Logo, S é da forma $\{i, i + 1, \dots, i + (k_n - 1)\}$ com $1 \leq i \leq 1 + L(k_n - 1)$. Nesse caso, $N_n = 1 + L(k_n - 1)$.
- *Estrelas*. Nesse exemplo, as componentes de X são as arestas de um grafo completo K_{k_n+1} com $k_n + 1$ vértices, logo $n = \binom{k_n+1}{2}$ e cada elemento S de \mathcal{C}_n é uma estrela de K_{k_n+1} , isto é, um subgrafo que contém todas as k_n arestas ligadas a um vértice fixo. Nesse caso, $N_n = k_n + 1$. Na figura 2.1 temos um exemplo de estrela no grafo completo K_9 .

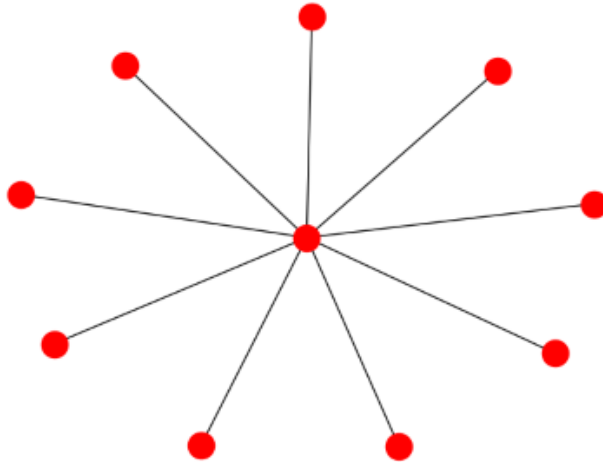


Figura 2.1: (Fonte : <https://www.johndcook.com/blog/2017/02/21/visualizing-graph-spectra-like-chemical-spectra/>.) Exemplo de estrela no grafo completo K_9 .

- *Árvores geradoras*. Nesse exemplo, as componentes de X são as arestas de um grafo completo K_{k_n+1} com $k_n + 1$ vértices, logo $n = \binom{k_n+1}{2}$ e cada elemento S de \mathcal{C}_n é uma árvore geradora de K_{k_n+1} . Uma árvore geradora de um grafo conexo é um subgrafo conexo sem ciclos e que

contém todas os vértices do grafo. Nesse caso, $N_n = (k_n + 1)^{k_n - 1}$ (veja Moon (1970)). Na figura 2.2 temos um exemplo de árvore geradora no grafo completo K_5 .

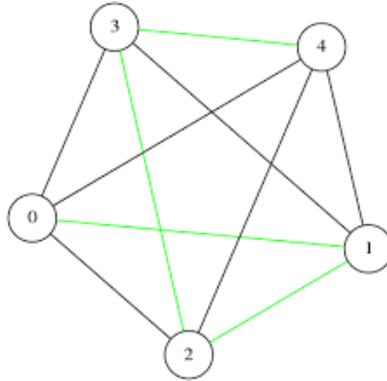


Figura 2.2: (Fonte : <http://www.martinbroadhurst.com/spanning-tree-of-a-graph-in-c.html>.) Exemplo de árvore geradora no grafo completo K_5 .

Um teste é uma função $f : \mathbb{R}^n \rightarrow \{0, 1\}$, onde $f(X) = 0$ indica que aceitamos a hipótese nula e caso contrário, a hipótese nula é rejeitada. Nessa dissertação, mediremos o desempenho de cada teste com base nos seguintes riscos:

- *Risco do pior caso possível.* Dada a classe \mathcal{C}_n , esse risco é definido para cada teste f por

$$R_{\max}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S\{f(X) = 0\}.$$

O menor risco do pior caso possível é chamado risco *minimax* e é definido como

$$R_{\max}^* := \inf_f R_{\max}(f),$$

onde o ínfimo é tomado sobre todos os testes possíveis.

- *Risco médio.* Dada a classe \mathcal{C}_n , esse risco é definido para cada teste f por

$$R(f) = \mathbb{P}_0\{f(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f(X) = 0\}, \quad (2.4)$$

onde $N_n := |\mathcal{C}_n|$ é a cardinalidade de \mathcal{C}_n . O menor risco médio é chamado risco de *Bayes* e é definido como

$$R^* := \inf_f R(f),$$

onde o ínfimo é tomado sobre todos os testes possíveis.

Observação 1. Note que o risco do pior caso possível e o risco médio dependem da medida de probabilidade \mathbb{P}_S que, por sua vez, depende do problema de detecção em questão. Dessa maneira, ambos os riscos também dependem do problema de detecção escolhido.

Dizemos que um teste h é *ótimo* com respeito ao risco de Bayes (respectivamente risco minimax) se $R^* = R(h)$ (respectivamente $R_{\max}^* = R_{\max}(h)$). Como veremos na Proposição 1 abaixo, o teste da razão de verossimilhança (TRV) é ótimo com respeito ao risco de Bayes, para ambos os problemas de detecção. Antes de enunciarmos formalmente esse resultado, vamos introduzir o TRV para ambos os problemas de detecção.

Seja $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ e denote

$$\phi_0(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}x^T \mathbf{I}_n x\right), \quad (2.5)$$

a densidade do vetor aleatório X sob \mathbb{P}_0 . A definição do TRV para os problemas de detecção de médias e correlações é dada abaixo.

- *Detecção de médias.* Para cada $S \in \mathcal{C}_n$, denotamos a densidade do vetor aleatório X sob \mathbb{P}_S como sendo

$$\phi_S(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}(x - \mu_S)^T \mathbf{I}_n (x - \mu_S)\right), \quad (2.6)$$

onde μ_S é definido em (2.1) e denotamos a razão de verossimilhança em x por

$$L(x) = \frac{(1/N_n) \sum_{S \in \mathcal{C}_n} \phi_S(x)}{\phi_0(x)} = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n x_S - k_n \mu_n^2 / 2}, \quad (2.7)$$

onde $x_S = \sum_{i \in S} x_i$. Assim, o teste da razão de verossimilhança é definido como

$$f^*(x) = \mathbb{1}_{\{L(x) > 1\}} = \begin{cases} 0, & \text{se } \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n x_S - k_n \mu_n^2 / 2} \leq 1, \\ 1, & \text{caso contrário.} \end{cases} \quad (2.8)$$

- *Detecção de correlações.* Para cada $S \in \mathcal{C}_n$, denotamos a densidade do vetor aleatório X sob \mathbb{P}_S como sendo

$$\phi_S(x) = (\det(A_S))^{-1/2} (2\pi)^{-n/2} \exp\left(-\frac{1}{2}x^T A_S^{-1} x\right), \quad (2.9)$$

onde A_S é definido em (2.3). Denotando

$$Z_S = \exp\left(\frac{1}{2}x^T(\mathbf{I}_n - A_S^{-1})x\right), \quad (2.10)$$

para todo $S \in \mathcal{C}_n$, definimos a razão de verossimilhança em x como

$$L(x) = \frac{(1/N_n) \sum_{S \in \mathcal{C}_n} \phi_S(x)}{\phi_0(x)} = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \frac{Z_S}{\sqrt{\det(A_S)}}.$$

Nesse caso, o teste da razão de verossimilhança é dado por

$$f^*(x) = \mathbb{1}_{\{L(x) > 1\}} = \begin{cases} 0, & \text{se } \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \frac{Z_S}{\sqrt{\det(A_S)}} \leq 1, \\ 1, & \text{caso contrário.} \end{cases} \quad (2.11)$$

O resultado abaixo afirma que o teste da razão de verossimilhança é ótimo com respeito ao risco de Bayes.

Proposição 1. *Para todo teste f , temos que $R(f) \geq R(f^*)$, onde f^* é o teste da razão de verossimilhança e $R(f)$ é o risco médio definido em (2.4).*

Durante a prova da Proposição 1, vamos fazer uso do seguinte lema.

Lema 1 (Neyman-Pearson). *Sejam \mathbb{Q}_0 e \mathbb{Q}_1 duas medidas de probabilidade em \mathbb{R}^n com densidades q_0 e q_1 em relação a uma medida σ -finita ν em \mathbb{R}^n . Defina $f_{TRV} : \mathbb{R}^n \rightarrow \{0, 1\}$ para todo $y \in \mathbb{R}^n$ por*

$$f_{TRV}(y) = \begin{cases} 1, & \text{se } q_0 \leq q_1, \\ 0, & \text{caso contrário.} \end{cases}$$

Então, para toda função $f : \mathbb{R}^n \rightarrow \{0, 1\}$ mensurável vale que

$$\mathbb{Q}_0\{y \in \mathbb{R}^n : f(y) = 1\} + \mathbb{Q}_1\{y \in \mathbb{R}^n : f(y) = 0\} \geq \int_{\mathbb{R}^n} \min(q_0(y), q_1(y)) d\nu(y).$$

A igualdade é satisfeita para o teste f_{TRV} .

A demonstração do Lema 1 é feita no Apêndice A.

Demonstração da Proposição 1. Defina $\mathbb{Q}_0 = \mathbb{P}_0 \circ X^{-1}$ e $\mathbb{Q}_S = \mathbb{P}_S \circ X^{-1}$ a distribuição de X sob \mathbb{P}_0 e \mathbb{P}_S , respectivamente. Veja que \mathbb{Q}_0 e \mathbb{Q}_S tem densidade ϕ_0 e ϕ_S em relação à medida de Lebesgue, onde ϕ_0 é definido em (2.5) e ϕ_S é definido em (2.6) se estivermos no problema de detecção

de médias e definido em (2.9) se estivermos no problema de detecção de correlações. Defina $Q_1 = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} Q_S$ e observe que para qualquer conjunto $A \in \mathbb{R}^n$ mensurável,

$$\begin{aligned} Q_1\{A\} &= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} Q_S\{A\} \\ &= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \int_A \phi_S(x) dx \\ &= \int_A \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \phi_S(x) dx \\ &= \int_A q_1(x) dx, \end{aligned}$$

onde $\phi_1(x) = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \phi_S(x)$.
Note que

$$\begin{aligned} R(f) &\stackrel{(i)}{=} \mathbb{P}_0\{f(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f(X) = 0\} \\ &= Q_0\{x \in \mathbb{R}^n : f(x) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} Q_S\{x \in \mathbb{R}^n : f(x) = 0\} \\ &= Q_0\{x \in \mathbb{R}^n : f(x) = 1\} + Q_1\{x \in \mathbb{R}^n : f(x) = 0\} \\ &\stackrel{(ii)}{\geq} Q_0\{x \in \mathbb{R}^n : f_{TRV}(x) = 1\} + Q_1\{x \in \mathbb{R}^n : f_{TRV}(x) = 0\} \\ &= \mathbb{P}_0\{f^*(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f^*(X) = 0\} \\ &\stackrel{(iii)}{=} R(f^*), \end{aligned}$$

onde em (i) e (iii) usamos a definição do risco médio, e em (ii) usamos o Lema 1 com ν igual a medida de Lebesgue em \mathbb{R}^n , $q_0 = \phi_0$ e $q_1 = \phi_1$. Para concluir a prova, note que $f_{TRV} = 1$, se e somente se, $\phi_0 \leq \phi_1 = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \phi_S$, que ocorre se e somente se, $f^* = 1$. □

Por definição $R(f) \leq R_{\max}(f)$, para todo teste f . Da Proposição 1 temos que $R(f^*) \leq R(f)$ para todo teste f . Logo, $R(f^*) \leq R_{\max}(f)$ para todo teste f e então segue que $R^* = R(f^*) \leq R_{\max}^*$.

Na proposição a seguir, reescrevemos o risco de Bayes em função da razão de verossimilhança e este resultado será usado nos Capítulos 3 e 4.

Proposição 2. *O risco de Bayes pode ser escrito como*

$$R^* = 1 - \frac{1}{2} \mathbb{E}_0 \left[|L(X) - 1| \right],$$

onde $L(X)$ depende do problema de detecção escolhido.

Demonstração. Sejam $\mathbb{P}_1 = (1/N_n) \sum_{S \in \mathcal{C}_n} \mathbb{P}_S$ e $\phi_1(x) = (1/N_n) \sum_{S \in \mathcal{C}_n} \phi_S(x)$, onde ϕ_S é definido em (2.6) se estivermos no problema de detecção de médias e definido em (2.9) se estivermos no problema de detecção de correlações. Escrevemos \mathbb{E}_1 para denotar a esperança sob a medida de probabilidade \mathbb{P}_1 . Considere ϕ_0 a função definida em (2.5) e note que o risco médio do teste ótimo pode ser escrito como

$$\begin{aligned} R^* &= R(f^*) \\ &\stackrel{(i)}{=} \mathbb{P}_0\{f^*(X) = 1\} + \mathbb{P}_1\{f^*(X) = 0\} \\ &= \mathbb{E}_0[\mathbb{1}_{\{f^*(X)=1\}}] + \mathbb{E}_1[\mathbb{1}_{\{f^*(X)=0\}}] \\ &\stackrel{(ii)}{=} \mathbb{E}_0[\mathbb{1}_{\{L(X) \geq 1\}}] + \mathbb{E}_1[1 - \mathbb{1}_{\{L(X) \geq 1\}}] \\ &\stackrel{(iii)}{=} \int_{\mathbb{R}^n} \phi_0(x) \mathbb{1}_{\{y:L(y) \geq 1\}}(x) dx + 1 - \int_{\mathbb{R}^n} \phi_1(x) \mathbb{1}_{\{y:L(y) \geq 1\}}(x) dx \\ &= 1 + \int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) \mathbb{1}_{\{y:L(y) \geq 1\}}(x) dx, \end{aligned} \tag{2.12}$$

onde em (i) usamos a definição de risco médio, em (ii) usamos que $f^*(x) = \mathbb{1}_{\{L(x) > 1\}}$ e em (iii) usamos o fato que X é vetor aleatório contínuo com densidade ϕ_0 sob \mathbb{P}_0 e com densidade ϕ_1 sob \mathbb{P}_1 . Por outro lado,

$$\begin{aligned} \mathbb{E}_0 \left[|L(X) - 1| \right] &= \int_{\mathbb{R}^n} \left| \frac{\phi_1(x)}{\phi_0(x)} - 1 \right| \phi_0(x) dx \\ &= \int_{\mathbb{R}^n} |\phi_1(x) - \phi_0(x)| dx \\ &= \int_{\mathbb{R}^n} (\phi_1(x) - \phi_0(x)) \mathbb{1}_{\{y:\phi_1(y) \geq \phi_0(y)\}}(x) dx \\ &\quad + \int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) \mathbb{1}_{\{y:\phi_0(y) \geq \phi_1(y)\}}(x) dx. \end{aligned} \tag{2.13}$$

Além disso, note que

$$\int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) \mathbb{1}_{\{y:\phi_0(y) \geq \phi_1(y)\}}(x) dx = \int_{\mathbb{R}^n} (\phi_1(x) - \phi_0(x)) \mathbb{1}_{\{y:\phi_1(y) \geq \phi_0(y)\}}(x) dx, \tag{2.14}$$

pois

$$\begin{aligned}
0 &= \int_{\mathbb{R}^n} \phi_0(x) dx - \int_{\mathbb{R}^n} \phi_1(x) dx \\
&= \int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) dx \\
&= \int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) \mathbb{1}_{\{y: \phi_1(y) \geq \phi_0(y)\}}(x) dx \\
&\quad + \int_{\mathbb{R}^n} (\phi_0(x) - \phi_1(x)) \mathbb{1}_{\{y: \phi_0(y) \geq \phi_1(y)\}}(x) dx.
\end{aligned}$$

Logo, substituindo a igualdade (2.14) na terceira igualdade de (2.13) temos que

$$\mathbb{E}_0 \left[|L(X) - 1| \right] = 2 \int_{\mathbb{R}^n} (\phi_1(x) - \phi_0(x)) \mathbb{1}_{\{y: \phi_1(y) \geq \phi_0(y)\}}(x) dx.$$

Substituindo a equação acima na sexta igualdade em (2.12), deduzimos que $R^* = 1 - \frac{1}{2} \mathbb{E}_0 \left[|L(X) - 1| \right]$, concluindo a prova do resultado. \square

Finalizamos esse capítulo apresentando cotas superior e inferior para R^* em termos da medida de afinidade de Bhattacharyya ([Bhattacharyya \(1946\)](#)) definida por

$$\rho = \frac{1}{2} \mathbb{E}_0 \left[\sqrt{L(X)} \right]. \quad (2.15)$$

Proposição 3 ([Devroye et al. \(1996\)](#)). *A seguinte cadeia de desigualdades é satisfeita:*

$$1 - \sqrt{1 - 4\rho^2} \leq R^* \leq 2\rho,$$

onde ρ é definido em (2.15) e depende do problema de detecção considerado.

Demonstração. Considerando a notação da prova da Proposição 2 e usando o Lema 1 temos que

$$\begin{aligned}
R^* &= \int_{\mathbb{R}^n} \min(\phi_0(x), \phi_1(x)) dx \\
&\stackrel{(i)}{=} \int_{\mathbb{R}^n} \min(1, L(x)) \phi_0(x) dx \\
&= \mathbb{E}_0 \left[\min(1, L(X)) \right] \\
&= \mathbb{E}_0 \left[\sqrt{\min(1, L(X))} \sqrt{\min(1, L(X))} \right] \\
&\stackrel{(ii)}{\leq} \mathbb{E}_0 \left[\sqrt{L(X)} \right] \\
&= 2\rho,
\end{aligned} \quad (2.16)$$

onde em (i) usamos que $L(x) = \phi_1(x)/\phi_0(x)$ e (ii) usamos que $\min(a, b) \leq a$ e $\min(a, b) \leq b$. Para finalizar a demonstração, resta mostrar que

$$1 - \sqrt{1 - 4\rho^2} \leq R^*,$$

e isso ocorre, se e somente se,

$$(1 - R^*)^2 \leq 1 - 4\rho^2.$$

Ou seja, quando

$$4\rho^2 \leq 1 - (1 - R^*)^2 = R^*(2 - R^*). \quad (2.17)$$

Para verificar a desigualdade (2.17) observe primeiro que

$$4\rho^2 = \left(\mathbb{E}_0 \left[\sqrt{L(X)} \right] \right)^2 = \left(\mathbb{E}_0 \left[\sqrt{\max(L(X), 1)} \sqrt{\min(L(X), 1)} \right] \right)^2.$$

Pela desigualdade de Cauchy-Schwarz, temos que

$$4\rho^2 \leq \mathbb{E}_0 \left[\max(L(X), 1) \right] \mathbb{E}_0 \left[\min(L(X), 1) \right] \stackrel{(ii)}{=} R^* \mathbb{E}_0 \left[\max(L(X), 1) \right], \quad (2.18)$$

onde em (ii) usamos a terceira igualdade em (2.16). Como

$$\mathbb{E}_0 \left[\max(L(X), 1) + \min(L(X), 1) \right] = \mathbb{E}_0 \left[L(X) + 1 \right] = 2,$$

segue que

$$\mathbb{E}_0 \left[\max(L(X), 1) \right] = 2 - \mathbb{E}_0 \left[\min(L(X), 1) \right] = 2 - R^*.$$

Substituindo a igualdade acima em (2.18) temos que

$$4\rho^2 \leq 1 - (1 - R^*)^2 = R^*(2 - R^*),$$

provando a desigualdade (2.17) e portanto o resultado segue. \square

Observação 2. Pela Proposição 3 temos que quando 2ρ é próximo de 1, o risco de Bayes R^* também é próximo 1. Por outro lado, quando 2ρ é próximo de zero, o risco de Bayes R^* também é próximo de zero.

Capítulo 3

Detecção de médias

3.1 Cotas superiores

Como mencionado no Capítulo 2, o teste f^* que minimiza o risco de Bayes é explicitamente determinado. No entanto, obter cotas superiores para $R(f^*)$ em função de n, k_n, μ_n e \mathcal{C}_n é em geral, difícil. Vimos pela Proposição 1 que para dar cotas superiores para $R(f^*)$, basta encontrar cotas superiores o risco médio de testes auxiliares. Nesta seção, discutiremos brevemente dois testes auxiliares, a saber, o teste baseado na soma e o teste baseado no máximo. Como veremos, em quase todos os exemplos das classes \mathcal{C}_n consideradas nessa dissertação, o risco de um desses testes auxiliares e o risco de Bayes têm o mesmo comportamento assintótico.

3.1.1 Teste baseado na soma

O teste é baseado no fato da soma das componentes de X ter distribuição normal com média zero sob \mathbb{P}_0 e ter distribuição normal com média $\mu_n k_n$ sob \mathbb{P}_S para todo $S \in \mathcal{C}_n$ (ver (2.1)). O teste baseado na soma é definido para todo $x \in \mathbb{R}^n$ por

$$f_{TS}(x) = \mathbb{1}_{\left\{\sum_{i=1}^n x_i > \mu_n k_n / 2\right\}}. \quad (3.1)$$

Proposição 4 (Addario-Berry et al. (2010)). *Seja $\delta_n \in (0, 1)$. O risco médio do teste baseado na soma f_{TS} satisfaz $R(f_{TS}) \leq \delta_n$, quando*

$$\mu_n \geq \sqrt{\frac{8n}{k_n^2} \log\left(\frac{2}{\delta_n}\right)}.$$

Usaremos na prova da Proposição 4 o seguinte lema.

Lema 2. *Seja Y uma variável aleatória tal que $Y \sim \mathcal{N}(\mu, \sigma^2)$. Para todo $t \in \mathbb{R}$ temos que $\mathbb{P}\{Y - \mu \geq t\} \leq \exp(-t^2/2\sigma^2)$.*

Demonstração. Veja que para todo $t \in \mathbb{R}$ e $\lambda > 0$, temos que

$$\begin{aligned} \mathbb{P}\{Y - \mu \geq t\} &= \mathbb{P}\{\exp(\lambda(Y - \mu)) \geq \exp(\lambda t)\} \\ &\stackrel{(i)}{\leq} \mathbb{E}[\exp(\lambda(X - \mu))]/e^{\lambda t} \\ &= M_{Y-\mu}(\lambda)/e^{\lambda t}, \end{aligned}$$

onde em (i) usamos a desigualdade de Markov. Denote por $M_Z(\lambda)$, a função geradora de momentos da variável aleatória Z .

Como $Y - \mu \sim \mathcal{N}(0, \sigma^2)$, logo $M_{Y-\mu}(\lambda) = e^{\sigma^2\lambda^2/2}$. Desse modo

$$\mathbb{P}\{Y - \mu \geq t\} \leq \inf_{\lambda > 0} \exp\left(\frac{\sigma^2\lambda^2}{2} - \lambda t\right) = \exp(-t^2/2\sigma^2),$$

concluindo a prova. □

Demonstração da Proposição 4. Veja que sob \mathbb{P}_0 , a soma $\sum_{i=1}^n X_i$ tem distribuição $\mathcal{N}(0, n)$ e sob \mathbb{P}_S , a soma $\sum_{i=1}^n X_i$ tem distribuição $\mathcal{N}(\mu_n k_n, n)$ para todo $S \in \mathcal{C}_n$. Assim, pela definição de f_{TS} temos que

$$\begin{aligned} R(f_{TS}) &= \mathbb{P}_0\{f_{TS}(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f_{TS}(X) = 0\} \\ &= \mathbb{P}_0\left\{\sum_{i=1}^n X_i > \mu_n k_n/2\right\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\left\{\sum_{i=1}^n X_i \leq \mu_n k_n/2\right\} \\ &= \mathbb{P}_0\left\{\sum_{i=1}^n X_i > \mu_n k_n/2\right\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\left\{\sum_{i=1}^n X_i - \mu_n k_n \leq -\mu_n k_n/2\right\} \\ &= \mathbb{P}_0\left\{\sum_{i=1}^n X_i > \mu_n k_n/2\right\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\left\{\sum_{i=1}^n X_i - \mu_n k_n \geq \mu_n k_n/2\right\}. \end{aligned} \tag{3.2}$$

Aplicando o Lema 2 com $\mu = \mu_n k_n$ e $\sigma^2 = n$ temos que

$$\begin{aligned} \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\left\{\sum_{i=1}^n X_i - \mu_n k_n \geq \mu_n k_n/2\right\} &\leq \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \exp(-\mu_n^2 k_n^2/8n) \\ &= \exp(-\mu_n^2 k_n^2/8n). \end{aligned} \tag{3.3}$$

Aplicando novamente o Lema 2 com $\mu = 0$ e $\sigma^2 = n$ para dar uma cota superior para a primeira probabilidade na última igualdade em (3.2), e juntando essa cota com a cota obtida em (3.3) temos que

$$R(f_{TS}) \leq 2 \exp(-\mu_n^2 k_n^2 / 8n).$$

Desse modo, temos que

$$R(f_{TS}) \leq \delta_n \text{ desde que } 2 \exp(-\mu_n^2 k_n^2 / 8n) \leq \delta_n.$$

Isolando μ_n na desigualdade acima temos que

$$\mu_n \geq \sqrt{\frac{8n}{k_n^2} \log\left(\frac{2}{\delta_n}\right)},$$

e o resultado segue. □

3.1.2 Teste baseado no máximo

Para cada $S \in \mathcal{C}_n$ definimos a variável aleatória $X_S := \sum_{i \in S} X_i$. Note que sob a hipótese alternativa existe $S \in \mathcal{C}_n$ tal que a variável aleatória X_S tem distribuição $\mathcal{N}(\mu_n k_n, k_n)$. Essa observação nos leva a definir o seguinte :

$$f_{TM}(x) = 1 \quad \text{se e somente se} \quad \max_{S \in \mathcal{C}_n} X_S \geq \frac{\mu_n k_n + \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{2}.$$

O teste f_{TM} é chamado teste baseado no máximo.

Proposição 5 (Addario-Berry et al. (2010)). *O risco médio do teste f_{TM} baseado no máximo satisfaz $R(f_{TM}) \leq \delta_n$ quando*

$$\mu_n \geq \frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{k_n} + 2\sqrt{\frac{2}{k_n} \log\left(\frac{2}{\delta_n}\right)}.$$

Para mostrar essa Proposição, vamos precisar do seguinte lema.

Lema 3. *Seja $Y = (Y_1, \dots, Y_n)$ vetor de n variáveis aleatórias independentes com distribuição normal padrão. Seja $g : \mathbb{R}^n \rightarrow \mathbb{R}$ função Lipschitz com constante de Lipschitz K (com respeito a distância Euclidiana). Então para todo $t > 0$,*

$$\mathbb{P}\{g(Y) - \mathbb{E}[g(Y)] \geq t\} \leq e^{-t^2/2K}.$$

A prova do Lema 3 é feita no Apêndice C.

Demonstração da Proposição 5. Veja que sob \mathbb{P}_0 , para cada $S \in \mathcal{C}_n$, $X_S \sim \mathcal{N}(0, k_n)$ onde $k_n = |S|$. Para usar o Lema 3, precisamos mostrar que a função $g(x) = g(x_1, \dots, x_n) = \max_{S \in \mathcal{C}_n} \sum_{i \in S} x_i$ é função Lipschitz e calcular a sua constante. Vamos mostrar que para todo $x, y \in \mathbb{R}^n$, temos que

$$|g(x) - g(y)| \leq \sqrt{k_n} \|x - y\|_2.$$

Seja $x_S = \sum_{i \in S} x_i$. Note que,

$$|g(x) - g(y)| = \left| \max_{S \in \mathcal{C}_n} x_S - \max_{S \in \mathcal{C}_n} y_S \right| \stackrel{(i)}{\leq} \max_{S \in \mathcal{C}_n} |x_S - y_S| = \max_{S \in \mathcal{C}_n} \left| \sum_{i \in S} (x_i - y_i) \right|.$$

Para ver que (i) vale, assuma primeiramente que $\max_{S \in \mathcal{C}_n} x_S \geq \max_{S \in \mathcal{C}_n} y_S$. Nesse caso, $|\max_{S \in \mathcal{C}_n} x_S - \max_{S \in \mathcal{C}_n} y_S| = \max_{S \in \mathcal{C}_n} x_S - \max_{S \in \mathcal{C}_n} y_S$. Seja S_j com $j \in \{1, \dots, N_n\}$ tal que $x_{S_j} = \max_{S \in \mathcal{C}_n} x_S$. Então usando essa notação, temos que

$$\max_{S \in \mathcal{C}_n} x_S - \max_{S \in \mathcal{C}_n} y_S = x_{S_j} - \max_{S \in \mathcal{C}_n} y_S \leq x_{S_j} - y_{S_j} \leq \max_{S \in \mathcal{C}_n} (x_S - y_S) \leq \max_{S \in \mathcal{C}_n} |x_S - y_S|.$$

O caso $\max_{S \in \mathcal{C}_n} x_S \leq \max_{S \in \mathcal{C}_n} y_S$ é simétrico. Agora, fixe $S \in \mathcal{C}_n$ qualquer e defina

$$(\mathbf{1}_S)_i = \begin{cases} 1, & \text{se } i \in S, \\ 0, & \text{caso contrário.} \end{cases}$$

Pela desigualdade de Cauchy-Schwarz temos que

$$\left| \sum_{i \in S} (x_i - y_i) \right| = |\langle x - y, \mathbf{1}_S \rangle| \leq \|\mathbf{1}_S\|_2 \|x - y\|_2 = \sqrt{k_n} \|x - y\|_2,$$

onde $\langle \cdot, \cdot \rangle$ denota o produto interno usual de \mathbb{R}^n . Logo, segue que

$$|g(x) - g(y)| \leq \max_{S \in \mathcal{C}_n} \left| \sum_{i \in S} (x_i - y_i) \right| \leq \max_{S \in \mathcal{C}_n} \sqrt{k_n} \|x - y\|_2 = \sqrt{k_n} \|x - y\|_2.$$

Aplicando o Lema 3 temos que para todo $t > 0$,

$$\mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X_S \geq \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] + t \right\} \leq \exp(-t^2/2k_n). \quad (3.4)$$

Assumindo que $\mu_n k_n > \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]$ e usando (3.4) com $t_n = (\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S])/2 > 0$ temos que

$$\mathbb{P}_0 \{f_{TM}(X) = 1\} = \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X_S \geq \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] + t_n \right\} \leq \exp(-t_n^2/2k_n).$$

Agora, note que

$$\begin{aligned}
\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \{f_{TM}(X) = 0\} &= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S' \in \mathcal{C}_n} X_{S'} \leq \frac{\mu_n k_n + \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{2} \right\} \\
&= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \bigcap_{S \in \mathcal{C}_n} \left\{ X_S \leq \frac{\mu_n k_n + \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{2} \right\} \right\} \\
&\leq \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ X_S \leq \frac{\mu_n k_n + \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{2} \right\} \\
&= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ X_S - \mu_n k_n \leq \frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] - \mu_n k_n}{2} \right\} \\
&= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ X_S - \mu_n k_n \geq \frac{\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{2} \right\} \\
&\stackrel{(ii)}{\leq} \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \exp \left(- \left(\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \right)^2 / 8k_n \right) \\
&= \exp \left(- \left(\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \right)^2 / 8k_n \right),
\end{aligned}$$

onde em (ii) usamos o Lema 2 com $\mu = \mu_n k_n$ e $\sigma^2 = k_n$. Portanto, pelas cotas superiores obtidas temos que se $\mu_n k_n > \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]$, então

$$\begin{aligned}
R(f_{TM}) &= \mathbb{P}_0 \{f_{TM}(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \{f_{TM}(X) = 0\} \\
&\leq 2 \exp \left(- \left(\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \right)^2 / 8k_n \right).
\end{aligned}$$

Para que $R(f_{TM}) \leq \delta_n$, precisamos que $\mu_n k_n > \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]$ e

$$2 \exp \left(- \left(\mu_n k_n - \mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \right)^2 / 8k_n \right) \leq \delta_n.$$

Isolando μ_n temos que

$$\mu_n \geq \frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{k_n} + 2 \sqrt{\frac{2}{k_n} \log \left(\frac{2}{\delta_n} \right)},$$

e o resultado segue. □

A Proposição 5 mostra que o risco médio do teste baseado no máximo é no máximo δ_n sempre que μ_n é pelo menos $\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]/k_n + 2\sqrt{2 \log(2/\delta_n)/k_n}$. Assim, para entender melhor o comportamento do teste baseado no máximo (e, assim, obter condições suficientes para que o risco do teste ótimo seja pequeno), é preciso entender o valor esperado de $\max_{S \in \mathcal{C}_n} X_S$ (sob \mathbb{P}_0). Como o máximo de processos Gaussianos tem sido estudado extensivamente, há muitos resultados diretamente aplicáveis disponíveis como o da seguinte proposição.

Proposição 6. *Seja $X = (X_1, \dots, X_n)$ vetor n -dimensional tal que sob \mathbb{P}_0 , X segue distribuição normal padrão multivariada. Defina $X_S = \sum_{i \in S} X_i$. Logo, vale que*

$$\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \leq \sqrt{2k_n \log N_n},$$

onde $|\mathcal{C}_n| = N_n$.

Demonstração. Pela desigualdade de Jensen, para todo t positivo temos que

$$\begin{aligned} \exp\left(t \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} X_S\right]\right) &\leq \mathbb{E}_0\left[\exp\left(t \max_{S \in \mathcal{C}_n} X_S\right)\right] \\ &= \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} \exp(tX_S)\right] \\ &\leq \sum_{S \in \mathcal{C}_n} \mathbb{E}_0[\exp(tX_S)] \\ &= N_n \exp(t^2 k_n / 2). \end{aligned}$$

Aplicando a função logarítmica e dividindo por t obtemos que

$$\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \leq \frac{\log N_n}{t} + \frac{tk_n}{2}.$$

Como a desigualdade acima vale para todo t positivo, vamos usá-la com t que minimize a função

$$g(t) = \frac{\log N_n}{t} + \frac{tk_n}{2}.$$

Para encontrar t que minimiza g vamos derivar g duas vezes, achar o ponto de inflexão da função g e verificar se de fato t é ponto de mínimo. Observe que

$$\frac{dg(t_*)}{dt} = -\frac{\log N_n}{t_*^2} + \frac{k_n}{2} = 0 \quad \text{se e somente se} \quad t_* = \sqrt{\frac{2 \log N_n}{k_n}},$$

e

$$\frac{d^2g(t_*)}{dt^2} = \frac{2 \log N_n}{(t_*)^3} = \frac{k_n \sqrt{k_n}}{\sqrt{2 \log N_n}} > 0.$$

Portanto, $t_* = \sqrt{\frac{2 \log N_n}{k_n}}$ é ponto de mínimo de g e segue que

$$\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S] \leq \frac{\log(N_n)}{t_*} + \frac{t_* k_n}{2} = \sqrt{2k_n \log N_n}.$$

□

Observação 3. Cotas superiores mais precisas podem ser obtidas por argumentos de acoplamento que podem ser encontrados em (Talagrand (2005)). Nessa dissertação, iremos usar inúmeras vezes o resultado da Proposição 6.

3.2 Cotas Inferiores

Nesta seção, obtemos cotas inferiores para R^* . Começamos com uma cota universal que garante que R^* é pelo menos δ_n , se μ_n for no máximo

$$\sqrt{\frac{4}{k_n} \log \left(1 + \frac{(1 - \delta_n)^2}{1 - (1 - \delta_n)^2} \right)},$$

independentemente da classe \mathcal{C}_n escolhida.

No resultado abaixo apresentamos uma cota universal para R^* .

3.2.1 Cota inferior universal

Proposição 7. *Seja $\delta_n \in (0, 1)$. Para todas as classes \mathcal{C}_n , $R^* \geq \delta_n$ sempre que*

$$\mu_n \leq \sqrt{\frac{4}{k_n} \log \left(1 + \frac{(1 - \delta_n)^2}{1 - (1 - \delta_n)^2} \right)}.$$

Para mostrar a Proposição 7, precisaremos do seguinte lema.

Lema 4. *Para o problema de detecção de médias, temos que a medida de afinidade de Bhattacharyya ρ definida em (2.15) satisfaz*

$$2\rho \geq e^{-\mu_n^2 k_n / 8}.$$

Demonstração. Usando a notação introduzida no Capítulo 2, em particular as definições (2.5) e (2.6), observe que pela desigualdade de Jensen,

$$2\rho = \mathbb{E}_0 \left[\sqrt{L(X)} \right] = \int_{\mathbb{R}^n} \sqrt{\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \phi_S(x) \phi_0(x)} dx \geq \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \int_{\mathbb{R}^n} \sqrt{\phi_S(x) \phi_0(x)} dx.$$

Vamos mostrar que para qualquer $S \in \mathcal{C}_n$,

$$\int_{\mathbb{R}^n} \sqrt{\phi_S(x) \phi_0(x)} dx = e^{-\mu_n^2 k_n / 8},$$

concluindo a prova do resultado.

Note que para todo $x \in \mathbb{R}^n$,

$$\begin{aligned} \sqrt{\phi_S(x) \phi_0(x)} &= \left(\frac{1}{(2\pi)^n} e^{-\sum_{i \in S} \frac{(x_i - \mu_n)^2}{2} - \sum_{i \notin S} \frac{x_i^2}{2} - \sum_{i=1}^n \frac{x_i^2}{2}} \right)^{1/2} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i \in S} \frac{(x_i - \mu_n)^2}{4} - \sum_{i \notin S} \frac{x_i^2}{4} - \sum_{i=1}^n \frac{x_i^2}{4}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i \in S} \frac{x_i^2 - 2x_i \mu_n + \mu_n^2}{4} - \sum_{i \notin S} \frac{x_i^2}{2} - \sum_{i \in S} \frac{x_i^2}{4}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i \in S} \frac{2x_i^2 - 4x_i \mu_n + 2\mu_n^2}{8} - \sum_{i \notin S} \frac{4x_i^2}{8} - \sum_{i \in S} \frac{2x_i^2}{8}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i \in S} \frac{4x_i^2 - 4x_i \mu_n + \mu_n^2}{8} - \sum_{i \notin S} \frac{4x_i^2}{8} - \sum_{i \in S} \frac{\mu_n^2}{8}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i \in S} \frac{(2x_i - \mu_n)^2}{8} - \sum_{i \notin S} \frac{4x_i^2}{8} - \frac{\mu_n^2 k_n}{8}} \\ &= \phi_{\mathcal{N}(a, \mathbf{I}_n)}(x) e^{-k_n \mu_n^2 / 8}, \end{aligned}$$

onde $a = (a_1, \dots, a_n)$ é um vetor de \mathbb{R}^n definido por

$$a_i = \begin{cases} \mu_n / 2, & \text{se } i \in S, \\ 0, & \text{caso contrário,} \end{cases}$$

\mathbf{I}_n é a matriz identidade de dimensão $n \times n$ e $\phi_{\mathcal{N}(a, \mathbf{I}_n)}(x)$ é a função de densidade de $\mathcal{N}(a, \mathbf{I}_n)$. Desse modo, ao integarmos nos valores possíveis do vetor x , segue que

$$\int_{\mathbb{R}^n} \sqrt{\phi_S(x) \phi_0(x)} dx = e^{-\mu_n^2 k_n / 8}.$$

□

Demonstração da Proposição 7. Pelo Lema 4 temos que

$$2\rho \geq e^{-\mu_n^2 k_n/8},$$

de modo que

$$1 - 4\rho^2 \leq 1 - e^{-\mu_n^2 k_n/4}. \quad (3.5)$$

Logo pela Proposição 3 temos que

$$R^* \geq \delta_n \quad \text{quando} \quad \delta_n \leq 1 - \sqrt{1 - 4\rho^2}.$$

Pela equação (3.5), temos que

$$\leq 1 - \sqrt{1 - e^{-\mu_n^2 k_n/4}} \leq 1 - \sqrt{1 - 4\rho^2}.$$

Logo, basta encontrar μ_n tal que

$$\delta_n \leq 1 - \sqrt{1 - e^{-\mu_n^2 k_n/4}},$$

o que ocorre se

$$\mu_n \leq \sqrt{\frac{4}{k_n} \log \left(1 + \frac{(1 - \delta_n)^2}{1 - (1 - \delta_n)^2} \right)},$$

concluindo a prova. \square

Concluimos da Proposição 7 que independente da classe \mathcal{C}_n escolhida, ao tomar $\delta_n \rightarrow 1$ quando $n \rightarrow \infty$ temos que $R^* \rightarrow 1$ quando $n \rightarrow \infty$, desde que

$$\mu_n \leq \sqrt{\frac{4}{k_n} \log \left(1 + \frac{(1 - \delta_n)^2}{1 - (1 - \delta_n)^2} \right)} \rightarrow 0,$$

quando $n \rightarrow \infty$. Portanto, é impossível detectar anomalias quando $\delta_n \rightarrow 1$ quando $n \rightarrow \infty$.

3.2.2 Cota inferior baseada em pares sobrepostos

No resultado abaixo, apresentamos uma cota inferior para o risco de Bayes baseada na função geradora de momentos de pares sobrepostos.

Proposição 8 (Arias-Castro et al. (2008)). *Sejam S e S' escolhidos uniformemente e de maneira independente em \mathcal{C}_n e seja $Z = |S \cap S'|$. Então,*

$$R^* \geq 1 - \frac{1}{2} \sqrt{\mathbb{E} [e^{\mu_n^2 Z}] - 1}.$$

Demonstração. Considerando a notação da prova da Proposição 2, note que

$$R^* \stackrel{(i)}{=} 1 - \frac{1}{2} \mathbb{E}_0 \left[|L(X) - 1| \right] \stackrel{(ii)}{\geq} 1 - \frac{1}{2} \sqrt{\mathbb{E}_0 \left[|L(X) - 1|^2 \right]},$$

onde em (i) usamos a Proposição 2 e em (ii) usamos a desigualdade de Jensen. Como

$$\mathbb{E}_0 [L(X)] = \int_{\mathbb{R}^n} \frac{\phi_1(x)}{\phi_0(x)} \phi_0(x) dx = \int_{\mathbb{R}^n} \phi_1(x) dx = 1,$$

logo

$$\mathbb{E}_0 [|L(X) - 1|^2] := \text{Var}_0(L(X)) = \mathbb{E}_0 [L^2(X)] - 1.$$

Lembre da equação (2.7) que,

$$L(X) = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S - k_n \mu_n^2 / 2}.$$

Logo,

$$\mathbb{E}_0 [L^2(X)] = \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} e^{-\mu_n^2 k_n} \mathbb{E}_0 \left[e^{\mu_n (X_S + X_{S'})} \right]. \quad (3.6)$$

Agora, veja que

$$\begin{aligned} \mathbb{E}_0 \left[e^{\mu_n (X_S + X_{S'})} \right] &= \mathbb{E}_0 \left[e^{\mu_n \sum_{i \in S \setminus S'} X_i} e^{\mu_n \sum_{i \in S' \setminus S} X_i} e^{2\mu_n \sum_{i \in S \cap S'} X_i} \right] \\ &= \mathbb{E}_0 \left[e^{\mu_n \sum_{i \in S \setminus S'} X_i} \right] \mathbb{E}_0 \left[e^{\mu_n \sum_{i \in S' \setminus S} X_i} \right] \mathbb{E}_0 \left[e^{2\mu_n \sum_{i \in S \cap S'} X_i} \right]. \end{aligned}$$

Observando que sob \mathbb{P}_0 , as variáveis aleatórias $\sum_{i \in S \setminus S'} X_i \sim \mathcal{N}(0, |S \setminus S'|)$, $\sum_{i \in S' \setminus S} X_i \sim \mathcal{N}(0, |S' \setminus S|)$ e $\sum_{i \in S \cap S'} X_i \sim \mathcal{N}(0, |S \cap S'|)$ segue que

$$\begin{aligned} \mathbb{E}_0 \left[e^{\mu_n (X_S + X_{S'})} \right] &= e^{\frac{\mu_n^2}{2} |S \setminus S'| + \frac{\mu_n^2}{2} |S' \setminus S| + 2\mu_n^2 |S \cap S'|} \\ &= e^{\frac{\mu_n^2}{2} (k_n - |S \cap S'|) + \frac{\mu_n^2}{2} (k_n - |S \cap S'|) + 2\mu_n^2 |S \cap S'|} \\ &= e^{\mu_n^2 k_n + \mu_n^2 |S \cap S'|}. \end{aligned} \quad (3.7)$$

Então, substituindo a última igualdade de (3.7) em (3.6) temos que

$$\mathbb{E}_0 [L^2(X)] = \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} e^{\mu_n^2 |S \cap S'|} = \mathbb{E} \left[e^{\mu_n^2 Z} \right].$$

Portanto

$$R^* \geq 1 - \frac{1}{2} \sqrt{\mathbb{E} \left[e^{\mu_n^2 Z} \right] - 1},$$

concluindo a prova. \square

A Proposição 8 reduz o problema de dar cotas inferiores para o risco de Bayes ao estudo da função geradora de momentos da sobreposição $|S \cap S'|$ entre dois elementos em \mathcal{C}_n escolhidos uniformemente e de maneira independente. Essa proposição é muito útil, como será ilustrado nas aplicações a seguir.

3.2.3 Cotas inferiores para classes simétricas

Como primeira aplicação da Proposição 8, vamos mostrar que a cota da Proposição 7 pode ser melhorada para toda classe \mathcal{C}_n suficientemente simétrica.

Proposição 9 (Addario-Berry et al. (2010)). *Fixe $\delta_n \in (0, 1)$. Sejam S, S' escolhidos uniformemente e de maneira independente em \mathcal{C}_n . Suponha que a classe \mathcal{C}_n satisfaça as seguintes condições de simetria:*

1. *A distribuição condicional de $Z = |S \cap S'|$ dado S' é idêntica para todos os valores de S' .*
2. *Para cada $S_0 \in \mathcal{C}_n$ e $i \in S_0$ fixados, $\mathbb{P}\{i \in S\} = \frac{k_n}{n}$.*

Então $R^* \geq \delta_n$ acontece para todo μ_n tal que

$$\mu_n \leq \sqrt{\frac{1}{k_n} \log \left(1 + \frac{4n(1 - \delta_n)^2}{k_n} \right)}.$$

Demonstração. Pela hipótese de simetria 1, veja que basta acharmos uma cota superior para $\mathbb{E} \left[e^{\mu_n^2 Z} \right] = \mathbb{E} \left[e^{\mu_n^2 Z} | S' \right]$ para um arbitrário $S' \in \mathcal{C}_n$. Assuma que $S' = \{1, \dots, k_n\}$, logo podemos escrever $Z = \sum_{i=1}^{k_n} \mathbb{1}_{\{i \in S\}}$. Veja que

$$\begin{aligned} \mathbb{E} \left[e^{\mu_n^2 Z} \right] &= \mathbb{E} \left[\prod_{i=1}^{k_n} e^{\mu_n^2 \mathbb{1}_{\{i \in S\}}} \right] \\ &\stackrel{(i)}{\leq} \prod_{i=1}^{k_n} \left(\mathbb{E} \left[e^{k_n \mu_n^2 \mathbb{1}_{\{i \in S\}}} \right] \right)^{1/k_n} \\ &\stackrel{(ii)}{=} \mathbb{E} \left[e^{k_n \mu_n^2 \mathbb{1}_{\{1 \in S\}}} \right] \\ &= e^{\mu_n^2 k_n} \mathbb{P}\{1 \in S\} + \mathbb{P}\{1 \notin S\} \\ &= \left(e^{\mu_n^2 k_n} - 1 \right) \frac{k_n}{n} + 1, \end{aligned} \tag{3.8}$$

onde em (i) usamos a desigualdade de Hölder e em (ii) usamos hipótese de simetria 2 da Proposição 9 . Usando a cota superior obtida em (3.8) e aplicando a Proposição 8 temos que $R^* \geq \delta_n$, desde que

$$1 - \frac{1}{2} \sqrt{\mathbb{E}[e^{\mu_n^2 Z}] - 1} \geq 1 - \frac{1}{2} \sqrt{\mathbb{E}\left[\left(e^{\mu_n^2 k_n} - 1\right) \frac{k_n}{n} + 1\right] - 1} \geq \delta_n.$$

Isolando μ_n na segunda desigualdade acima temos que $R^* \geq \delta_n$ desde que

$$\mu_n \leq \sqrt{\frac{1}{k_n} \log\left(1 + \frac{4n(1 - \delta_n)^2}{k_n}\right)},$$

concluindo a prova. \square

Corolário 1 (Addario-Berry et al. (2010)). *Seja $\delta_n \in (0, 1)$. Assuma que a classe \mathcal{C}_n é simétrica com respeito às hipóteses presentes na Proposição 9 e que contenha no máximo n^α elementos onde $\alpha > 0$. Então, $R^* \geq 1 - \delta_n$ para todo μ_n tal que*

$$\mu_n \leq \sqrt{\frac{1}{k_n} \log\left(1 + \frac{4n\delta_n^2}{k_n}\right)},$$

e $R^* \leq \delta_n$ para todo μ_n tal que

$$\mu_n \geq \sqrt{\frac{2\alpha}{k_n} \log n} + 2\sqrt{\frac{2 \log(2/\delta_n)}{k_n}}.$$

Demonstração. A primeira afirmação segue da Proposição 9 substituindo δ_n por $1 - \delta_n$. Já a segunda afirmação, segue da Proposição 5 e do fato que

$$\mathbb{E}_0[\max_{S \in \mathcal{C}} X_S] \stackrel{(i)}{\leq} \sqrt{2k_n \log N_n} \stackrel{(ii)}{\leq} \sqrt{2k_n \alpha \log n}$$

onde em (i) usamos a Proposição 6 e em (ii) usamos que $N_n \leq n^\alpha$. \square

Suponha que $k_n = n^\beta$ para algum $\beta \in (0, 1)$ tal que $1 - \beta < 2\alpha$. Suponha que $\delta_n \rightarrow 0$ de modo que $\log(1/\delta_n) \ll \log n$. Nesse caso, veja que

$$\begin{aligned} \sqrt{\frac{1}{k_n} \log\left(\frac{4n\delta_n^2}{k_n}\right)} &= \sqrt{\frac{1}{n^\beta} \log(4n^{1-\beta}\delta_n^2)} \\ &= \sqrt{\frac{1}{n^\beta} \log(n^{1-\beta}) + \frac{\log 4}{n^\beta} - \frac{2 \log(1/\delta_n)}{n^\beta}} \\ &= \sqrt{\frac{(1-\beta)}{n^\beta} \log n \left(1 + \frac{\log 4}{(1-\beta) \log n} - \frac{2 \log(1/\delta_n)}{(1-\beta) \log n}\right)} \\ &\approx \sqrt{\frac{(1-\beta)}{n^\beta} \log n}, \end{aligned}$$

pois $\log 4 \ll \log n$ e $\log(1/\delta_n) \ll \log n$. Agora, observe que

$$\begin{aligned} \sqrt{\frac{2\alpha}{k_n} \log n} + 2\sqrt{\frac{2 \log(2/\delta_n)}{k_n}} &= \sqrt{\frac{2\alpha}{n^\beta} \log n} + 2\sqrt{\frac{2 \log(2/\delta_n)}{n^\beta}} \\ &= \sqrt{\frac{2\alpha}{n^\beta} \log n} \left(1 + 2\sqrt{\frac{\log(2/\delta_n)}{\alpha \log n}} \right) \\ &\approx \sqrt{\frac{2\alpha}{n^\beta} \log n}, \end{aligned}$$

pois $\log(1/\delta_n) \ll \log n$. Como $\sqrt{\frac{1}{k_n} \log\left(\frac{4n\delta_n^2}{k_n}\right)} \leq \sqrt{\frac{1}{k_n} \log\left(1 + \frac{4n\delta_n^2}{k_n}\right)}$, logo pelo Corolário 1 temos que :

- Se $\mu_n \leq \sqrt{\frac{(1-\beta)}{n^\beta} \log n}$, então $R^* \rightarrow 1$ quando $n \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.
- Se $\mu_n \geq \sqrt{\frac{2\alpha}{n^\beta} \log n}$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.

Nesse caso, sob as hipóteses $k_n = n^\beta$ para algum $\beta \in (0, 1)$ tal que $1 - \beta < 2\alpha$ e $\delta_n \rightarrow 0$ de modo que $\log(1/\delta_n) \ll \log n$, a ordem de magnitude assintótica para o valor crítico de μ_n é $\sqrt{(\log n)/k_n}$.

A cota na Proposição 9 pode ser melhorada sob a condição adicional de correlação negativa.

Definição 1. Para quaisquer $x, y \in \mathbb{R}^d$, dizemos que $x \leq_d y$ se $x_i \leq y_i$ para todo $1 \leq i \leq d$. Seja $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$. Dizemos que f é não decrescente se $f(x) \leq f(y)$ para $x, y \in U$ tal que $x \leq_d y$

Definição 2. Sejam Y_1, \dots, Y_n variáveis aleatórias onde $Y_i \subset U_i \subset \mathbb{R}$ para todo $1 \leq i \leq n$. A coleção Y_1, \dots, Y_n é dita negativamente correlacionada se para qualquer par de conjuntos disjuntos $I, J \subset \{1, \dots, n\}$ e funções não decrescentes $f : \prod_{i \in I} U_i \rightarrow \mathbb{R}$ e $g : \prod_{j \in J} U_j \rightarrow \mathbb{R}$, temos

$$\mathbb{E} [f(Y_i, i \in I)g(Y_j, j \in J)] \leq \mathbb{E} [f(Y_i, i \in I)] \mathbb{E} [g(Y_j, j \in J)].$$

Proposição 10 (Addario-Berry et al. (2010)). *Seja $\delta_n \in (0, 1)$ e suponha que a classe \mathcal{C}_n satisfaz as condições da Proposição 9. Suponha que $S' = \{1, \dots, k_n\} \in \mathcal{C}_n$. Seja S escolhido ao acaso em \mathcal{C}_n . Se as variáveis aleatórias*

$\mathbb{1}_{\{1 \in S\}}, \dots, \mathbb{1}_{\{k_n \in S\}}$ são negativamente correlacionadas, então $R^* \geq \delta_n$ para todo μ_n tal que

$$\mu_n \leq \sqrt{\log \left(1 + \frac{n \log(1 + 4(1 - \delta_n)^2)}{k_n^2} \right)}.$$

Demonstração. Vamos proceder de maneira análoga a demonstração da Proposição 9. Note que

$$\begin{aligned} \mathbb{E} \left[e^{\mu_n^2 Z} \right] &= \mathbb{E} \left[\prod_{i=1}^{k_n} e^{\mu_n^2 \mathbb{1}_{\{i \in S\}}} \right] \\ &\stackrel{(i)}{\leq} \prod_{i=1}^{k_n} \mathbb{E} \left[e^{\mu_n^2 \mathbb{1}_{\{i \in S\}}} \right] \\ &\stackrel{(ii)}{=} \left(\left(e^{\mu_n^2} - 1 \right) \frac{k_n}{n} + 1 \right)^{k_n}, \end{aligned} \tag{3.9}$$

onde em (i) usamos a hipótese que as variáveis aleatórias $\mathbb{1}_{\{1 \in S\}}, \dots, \mathbb{1}_{\{k_n \in S\}}$ são negativamente correlacionadas e em (ii) usamos a hipótese 2 da Proposição 9. Por outro lado, como estamos nas condições da Proposição 9, $R^* \geq \delta_n$ quando

$$1 - \frac{1}{2} \sqrt{\mathbb{E} \left[e^{\mu_n^2 Z} \right] - 1} \geq \delta_n,$$

isto é, se

$$\mathbb{E} \left[e^{\mu_n^2 Z} \right] \leq 1 + 4(1 - \delta_n)^2. \tag{3.10}$$

Assim, por (3.9) e (3.10), queremos achar μ_n tal que

$$\mathbb{E} \left[e^{\mu_n^2 Z} \right] \leq \left(\left(e^{\mu_n^2} - 1 \right) \frac{k_n}{n} + 1 \right)^{k_n} \leq 1 + 4(1 - \delta_n)^2.$$

Isolando μ_n na segunda desigualdade acima temos que

$$\mu_n \leq \sqrt{\log \left(1 + \frac{n}{k_n} \left((1 + 4(1 - \delta_n)^2)^{1/k_n} - 1 \right) \right)}. \tag{3.11}$$

Usando a desigualdade $e^y \geq 1 + y$ com $y = k_n^{-1} \log(1 + 4(1 - \delta_n)^2)$ obtemos que

$$(1 + 4(1 - \delta_n)^2)^{1/k_n} - 1 \geq \frac{\log((1 + 4(1 - \delta_n)^2))}{k_n}.$$

Após substituir a inequação acima em (3.11), obtemos que $R^* \geq \delta_n$ se

$$\mu_n \leq \sqrt{\log \left(1 + \frac{n \log(1 + 4(1 - \delta_n)^2)}{k_n^2} \right)},$$

concluindo a prova. \square

3.3 Exemplos

Nesta seção, vamos considerar exemplos concretos e trabalhar com as cotas superiores e inferiores obtidas, respectivamente, na Seção 3.1 e na Seção 3.2, de modo a determinar quando é possível ou impossível detectar anomalias.

3.3.1 Conjuntos disjuntos

Considere o exemplo em que todos os conjuntos $S \in \mathcal{C}_n$ são disjuntos e dessa forma, $k_n N_n \leq n$. Fixe $\delta_n \in (0, 1)$. Veja que sob \mathbb{P}_0 , a variável aleatória $X_S = \sum_{i \in S} X_i \sim \mathcal{N}(0, k_n)$ para todo $S \in \mathcal{C}_n$. Sabemos da Proposição 5 que $R^* \leq \delta_n$ sempre que

$$\mu_n \geq \frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{k_n} + 2\sqrt{\frac{2 \log(2/\delta_n)}{k_n}}.$$

Da Proposição 6 temos que $\mathbb{E}_0[\max_{S \in \mathcal{C}} X_S] \leq \sqrt{2k_n \log N_n}$, logo $R^* \leq \delta_n$ quando

$$\mu_n \geq \sqrt{\frac{2 \log N_n}{k_n}} + 2\sqrt{\frac{2 \log(2/\delta_n)}{k_n}}. \quad (3.12)$$

Para obter a cota inferior do risco de Bayes, vamos usar a Proposição 8. Nesse caso, temos que

$$Z = \begin{cases} k_n, & \text{com probabilidade } \frac{1}{N_n}, \\ 0, & \text{com probabilidade } 1 - \frac{1}{N_n}. \end{cases}$$

Desse modo,

$$\mathbb{E} \left[e^{\mu_n^2 Z} \right] - 1 = \frac{1}{N_n} \left(e^{\mu_n^2 k_n} - 1 \right) \leq \frac{1}{N_n} e^{\mu_n^2 k_n}, \quad (3.13)$$

e pela Proposição 8 temos que $R^* \geq \delta_n$ quando

$$1 - \frac{1}{2} \sqrt{\frac{1}{N_n} e^{\mu_n^2 k_n}} \geq \delta_n.$$

Isolando μ_n na desigualdade acima obtemos que $R^* \geq \delta_n$ se

$$\mu_n \leq \sqrt{\frac{\log(4N_n(1 - \delta_n)^2)}{k_n}}. \quad (3.14)$$

Suponha que $\delta_n \rightarrow 0$ de modo que $\log(1/\delta_n) \ll \log N_n$. Nesse caso temos que

$$\begin{aligned} \sqrt{\frac{2 \log N_n}{k_n}} + 2 \sqrt{\frac{2 \log(2/\delta_n)}{k_n}} &= \sqrt{\frac{2 \log N_n}{k_n}} \left(1 + 2 \sqrt{\frac{\log(2/\delta_n)}{\log N_n}} \right) \\ &\approx \sqrt{\frac{2 \log N_n}{k_n}}, \end{aligned}$$

pois $\log(1/\delta_n) \ll \log N_n$. Logo, assintoticamente em (3.12)

$$\mu_n \geq \sqrt{2 \log(N_n)/k_n} \quad \text{implica que} \quad R^* \rightarrow 0 \quad \text{quando} \quad n \rightarrow \infty.$$

Por outro lado, veja que

$$\begin{aligned} \sqrt{\frac{\log(4N_n\delta_n^2)}{k_n}} &= \sqrt{\frac{\log N_n}{k_n} + \frac{\log 4}{k_n} - \frac{2 \log \delta_n}{k_n}} \\ &= \sqrt{\frac{\log N_n}{k_n} \left(1 + \frac{\log 4}{\log N_n} - \frac{2 \log \delta_n}{\log N_n} \right)} \\ &\approx \sqrt{\frac{\log N_n}{k_n}} \end{aligned}$$

pois $\log(1/\delta_n) \ll \log N_n$. Logo, assintoticamente em (3.14)

$$\mu_n \leq \sqrt{\log(N_n)/k_n} \quad \text{implica que} \quad R^* \rightarrow 1 \quad \text{quando} \quad n \rightarrow \infty.$$

Portanto, se $\log(1/\delta_n) \ll \log N_n$ com $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$ temos que:

- Se $\mu_n \geq \sqrt{2 \log(N_n)/k_n}$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.
- Se $\mu_n \leq \sqrt{\log(N_n)/k_n}$, então $R^* \rightarrow 1$ quando $n \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.

Nesse caso, a ordem de magnitude assintótica para o valor crítico de μ_n é $\sqrt{(\log N_n)/k_n}$.

3.3.2 k_n -conjuntos

Considere o exemplo em que \mathcal{C}_n contém todos os conjuntos $S \subset \{1, \dots, n\}$ que têm tamanho k_n . Nesse caso, $N_n = \binom{n}{k_n}$. Fixe $\delta_n \in (0, 1)$. Vamos aplicar a Proposição 10 e para isso precisamos verificar as suas hipóteses.

Seja $S' = S_0 = \{i_1, \dots, i_{k_n}\}$. Note que para todo $x \in \{0, \dots, k_n\}$ temos que,

$$\begin{aligned} \mathbb{P}\{Z = x | S' = S_0\} &= \mathbb{P}\{|S \cap S'| = x | S' = S_0\} \\ &= \mathbb{P}\{|S \cap S_0| = x | S' = S_0\} \\ &\stackrel{(i)}{=} \mathbb{P}\{|S \cap S_0| = x\} \\ &= \frac{\binom{k_n}{x} \binom{n-k_n}{k_n-x}}{N_n}, \end{aligned}$$

onde em (i) usamos que S e S' são independentes. Veja que a probabilidade acima não depende da escolha feita para S_0 . Portanto, a distribuição condicional de $|S \cap S'|$ dado S' é idêntica para todos os valores de S' . Além disso, para cada $i \in S_0$ veja que

$$\mathbb{P}\{i \in S\} = \frac{\binom{n-1}{k_n-1}}{\binom{n}{k_n}} = \frac{k_n}{n}.$$

Falta verificar que as variáveis aleatórias $\mathbb{1}_{\{1 \in S\}}, \dots, \mathbb{1}_{\{k_n \in S\}}$ são negativamente correlacionadas. Dizemos que σ é uma permutação do conjunto $\{1, \dots, n\}$ se $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ é uma função bijetora.

Definição 3. Seja $z = (z_1, \dots, z_n) \in \mathbb{R}^n$. Dizemos que o vetor aleatório $Z = (Z_1, \dots, Z_n)$ tem distribuição de permutação se:

1. $Z \in \{z_\sigma : \sigma \text{ é permutação de } \{1, \dots, n\}\}$, onde $z_\sigma = (z_{\sigma(1)}, \dots, z_{\sigma(n)})$.
2. Para toda σ permutação de $\{1, \dots, n\}$, temos que $\mathbb{P}\{Z = z_\sigma\} = \frac{1}{n!}$.

Teorema 1. *Todo vetor aleatório com distribuição de permutação é negativamente correlacionado.*

A prova do teorema acima não será apresentada nessa dissertação. Pode-se encontrar a demonstração em (Joag-Dev and Proschan (1983)), ver Teorema 2.11).

Veja que escolher uniformemente $S \in \mathcal{C}_n$ é equivalente a fazer k_n extrações sem reposição de uma urna com bolas numeradas de 1 à n . Defina

$$Z_i = \begin{cases} 1, & \text{se a bola } i \text{ foi escolhida,} \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$z = \underbrace{(1, \dots, 1)}_{k_n \text{ vezes}}, \underbrace{(0, \dots, 0)}_{n-k_n \text{ vezes}}.$$

Segue que o vetor aleatório $Z = (Z_1, \dots, Z_n)$ tem distribuição de permutação e então pelo Teorema 1 temos que Z é negativamente correlacionado.

Proposição 11. *Funções não decrescentes definidas em subconjuntos disjuntos de um conjunto de variáveis aleatórias negativamente correlacionadas também são negativamente correlacionadas.*

A prova da proposição acima não será apresentada e pode ser encontrada em (Joag-Dev and Proschan (1983), ver Propriedade 6). Definindo $f_i : \{0, 1\}^n \rightarrow \{0, 1\}$ por $f_i(x) = \mathbb{1}_{\{x_i=1\}}$, para todo $1 \leq i \leq k_n$ temos que f_i é não decrescente e $f_i(Z) = \mathbb{1}_{\{Z_i=1\}}$. Desse modo, pela Proposição 11 as variáveis aleatórias $\mathbb{1}_{\{Z_1=1\}}, \dots, \mathbb{1}_{\{Z_{k_n}=1\}}$ são negativamente correlacionadas. Como $\mathbb{1}_{\{Z_i=1\}} = \mathbb{1}_{\{i \in S\}}$, segue que as variáveis aleatórias $\mathbb{1}_{\{1 \in S\}}, \dots, \mathbb{1}_{\{k_n \in S\}}$ são negativamente correlacionadas.

Portanto, pela Proposição 10 segue que $R^* \geq \delta_n$ desde que

$$\mu_n \leq \sqrt{\log \left(1 + \frac{n \log(1 + 4(1 - \delta_n)^2)}{k_n^2} \right)}. \quad (3.15)$$

Por outro lado, pela Proposição 5 temos que $R^* \leq \delta_n$ desde que

$$\mu_n \geq \frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{k_n} + 2\sqrt{\frac{2}{k_n} \log \left(\frac{2}{\delta_n} \right)}. \quad (3.16)$$

Como $N_n = \binom{n}{k_n}$, logo pela Proposição 6 temos que

$$\frac{\mathbb{E}_0[\max_{S \in \mathcal{C}_n} X_S]}{k_n} \leq \sqrt{\frac{2 \log N_n}{k_n}} = \sqrt{\frac{2}{k_n} \log \binom{n}{k_n}}. \quad (3.17)$$

Agora, vamos dar cota superior para $\binom{n}{k_n}$. Para isso, precisamos do seguinte lema.

Lema 5. *Para quaisquer $1 \leq k_n \leq n$, vale que*

$$\binom{n}{k_n} \leq \left(\frac{en}{k_n} \right)^{k_n}.$$

Demonstração. Vamos mostrar o resultado por indução. Veja que para $k_n = 1$, a desigualdade vale pois para $n \geq 1$,

$$\binom{n}{1} = n \leq en = \left(\frac{en}{1}\right)^1.$$

Suponha que a desigualdade vale para $k_n \leq n - 1$. Mostraremos que valerá para o seu sucessor. Observe que

$$\begin{aligned} \binom{n}{k_n + 1} &= \frac{n!}{(k_n + 1)!(n - (k_n + 1))!} \\ &= \frac{n!(n - k_n)}{(k_n + 1)k_n!(n - k_n)!} \\ &= \binom{n}{k_n} \frac{n - k_n}{k_n + 1} \\ &\stackrel{(i)}{\leq} \left(\frac{en}{k_n}\right)^{k_n} \frac{n - k_n}{k_n + 1} \\ &\leq \left(\frac{en}{k_n}\right)^{k_n} \frac{n}{k_n + 1} \\ &= \frac{e^{k_n} n^{k_n+1}}{(k_n + 1)^{k_n+1}} \left(\frac{k_n + 1}{k_n}\right)^{k_n} \\ &= \frac{e^{k_n} n^{k_n+1}}{(k_n + 1)^{k_n+1}} \left(1 + \frac{1}{k_n}\right)^{k_n} \\ &\stackrel{(ii)}{\leq} \frac{e^{k_n+1} n^{k_n+1}}{(k_n + 1)^{k_n+1}}, \end{aligned}$$

onde em (i) usamos a hipótese de indução e em (ii) usamos que $e^x \geq 1 + x$ para x positivo. \square

Aplicando o Lema 5 em (3.17) temos que

$$\frac{\mathbb{E}_0[\max_{S \in \mathcal{C}} X_S]}{k_n} \leq \sqrt{\frac{2}{k_n} \log \binom{n}{k_n}} \leq \sqrt{2 \log \left(\frac{en}{k_n}\right)},$$

e assim por (3.16), concluímos que $R^* \leq \delta_n$ desde que

$$\mu_n \geq \sqrt{2 \log \left(\frac{en}{k_n}\right)} + 2\sqrt{\frac{2}{k_n} \log \left(\frac{2}{\delta_n}\right)}. \quad (3.18)$$

A seguir vamos estudar a ordem de magnitude do valor crítico μ_n para os casos em que $k_n = n^{(1-\epsilon)/2}$ e $k_n = n^{(1+\epsilon)/2}$, para algum $\epsilon > 0$ fixado.

Caso $k_n = n^{(1-\epsilon)/2}$. Nesse caso, por (3.15) temos que $R^* \geq 1 - \delta_n$ quando

$$\mu_n \leq \sqrt{\log(1 + n^\epsilon \log(1 + 4\delta_n^2))}, \quad (3.19)$$

e por (3.18) temos que $R^* \leq \delta_n$ quando

$$\mu_n \geq \sqrt{2 \log(en^{(1+\epsilon)/2})} + 2\sqrt{\frac{2 \log(2/\delta_n)}{n^{(1-\epsilon)/2}}}. \quad (3.20)$$

Suponha que $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$ de modo que $\log(\log(1 + 4\delta_n^2)) \ll \log n$ e $\log(1/\delta_n) \ll n^{(1-\epsilon)/2}$. Nesse caso, quando $n \rightarrow \infty$ temos que

$$\begin{aligned} \log(n^\epsilon \log(1 + 4\delta_n^2)) &= \epsilon \log n + \log(\log(1 + 4\delta_n^2)) \\ &= \log n \left(\epsilon + \frac{\log(\log(1 + 4\delta_n^2))}{\log n} \right) \\ &\approx \epsilon \log n, \end{aligned}$$

e

$$\begin{aligned} \sqrt{2 \log(en^{(1+\epsilon)/2})} + 2\sqrt{\frac{2 \log(1/\delta_n)}{n^{(1-\epsilon)/2}}} &= \sqrt{2 + (1 + \epsilon) \log n} + 2\sqrt{\frac{2 \log(1/\delta_n)}{n^{(1-\epsilon)/2}}} \\ &\approx \sqrt{2 + (1 + \epsilon) \log n}. \end{aligned}$$

Como $\log(n^\epsilon \log(1 + 4\delta_n^2)) \leq \log(1 + n^\epsilon \log(1 + 4\delta_n^2))$, logo de (3.19) e (3.20) temos que :

- Se $\mu_n \geq \sqrt{2 + (1 + \epsilon) \log n}$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.
- Se $\mu_n \leq \sqrt{\epsilon \log n}$, então $R^* \rightarrow 1$ quando $n \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.

Portanto, a ordem de magnitude assintótica nesse caso para o valor crítico de μ_n é $\sqrt{\log n}$.

Caso $k_n = n^{(1+\epsilon)/2}$. Nesse caso, por (3.15) temos que $R^* \geq 1 - \delta_n$ quando

$$\mu_n \leq \sqrt{\log\left(1 + \frac{1}{n^\epsilon} \log(1 + 4\delta_n^2)\right)},$$

e pela Proposição 4 temos que $R^* \leq \delta_n$ desde que

$$\mu_n \geq \sqrt{8n^{-\epsilon} \log(2/\delta_n)}.$$

Suponha que $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$. Como $\log(1 + 4\delta_n^2)/n^\epsilon \rightarrow 0$ quando $n \rightarrow \infty$ e já que $\log(1 + x)/x \rightarrow 1$ quando $x \rightarrow 0$, logo

$$\sqrt{\log\left(1 + \frac{1}{n^\epsilon} \log(1 + 4\delta_n^2)\right)} \approx \sqrt{\frac{1}{n^\epsilon} \log(1 + 4\delta_n^2)} \approx \sqrt{\frac{4\delta_n^2}{n^\epsilon}},$$

pois $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$. Desse modo, $R^* \geq 1 - \delta_n$ desde que

$$\mu_n \leq \sqrt{\frac{4\delta_n^2}{n^\epsilon}},$$

e $R^* \leq \delta_n$ desde que

$$\mu_n \geq \sqrt{8n^{-\epsilon} \log(2/\delta_n)}.$$

Portanto, para todo $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$ temos que:

- Se $\mu_n \leq \sqrt{\frac{4\delta_n^2}{n^\epsilon}}$, então $R^* \rightarrow 1$ quando $n \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.
- Se $\mu_n \geq \sqrt{8n^{-\epsilon} \log(2/\delta_n)}$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.

Observe que para qualquer $\delta_n \rightarrow 0$ quando $n \rightarrow \infty$ temos que

$$\sqrt{\frac{4\delta_n^2}{n^\epsilon}} \ll \sqrt{8n^{-\epsilon} \log(2/\delta_n)}.$$

Em particular, supondo que $\log(1/\delta_n) \ll n^\epsilon$ temos que

$$\sqrt{\frac{4\delta_n^2}{n^\epsilon}} \ll \sqrt{8n^{-\epsilon} \log(2/\delta_n)} \rightarrow 0,$$

de forma que se $\liminf_{n \rightarrow \infty} \mu_n > 0$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$.

3.3.3 Estrelas

Considere a rede de m vértices nos quais cada par de vértices interage. Queremos testar se existe um vértice na rede cuja interação é diferente do resto. Essa situação pode ser modelada considerando a classe das *estrelas*.

Uma estrela é um subgrafo do grafo completo K_m o qual contém todas as $k_n = m - 1$ arestas ligadas a um vértice fixo. Considere a classe \mathcal{C}_m de todas as estrelas do grafo completo K_m . Nesse exemplo, $n = \binom{m}{2}$ e $N_n = m$. Vamos aplicar a Proposição 9 e para isso precisamos verificar as suas hipóteses.

Seja $S' = S_0 = \{i_1, \dots, i_{k_n}\}$. Note que, para $x \in \{1, m-1\}$ temos que

$$\begin{aligned} \mathbb{P}\{Z = x | S' = S_0\} &= \mathbb{P}\{|S \cap S'| = x | S' = S_0\} \\ &= \mathbb{P}\{|S \cap S_0| = x | S' = S_0\} \\ &\stackrel{(i)}{=} \mathbb{P}\{|S \cap S_0| = x\} \\ &= \begin{cases} 1 - \frac{1}{m}, & \text{se } x = 1, \\ \frac{1}{m}, & \text{se } x = m-1. \end{cases} \end{aligned}$$

onde em (i) usamos que S e S' são independentes. Veja que a probabilidade acima não depende da escolha feita para S_0 . Portanto, a distribuição condicional de $|S \cap S'|$ dado S' é idêntica para todos os valores de S' . Além disso, para uma aresta i pertencer a uma estrela, um dos vértices tem que ser o “centro” da estrela. Como temos m vértices, segue que

$$\mathbb{P}\{i \in S\} = \frac{2}{m} = \frac{m-1}{\binom{m}{2}} = \frac{k_n}{n}.$$

Logo, as condições de simetria presentes na Proposição 9 são satisfeitas e segue que $R^* \geq 1 - \delta_m$ quando

$$\mu_m \leq \sqrt{\frac{1}{m-1} \log \left(1 + \frac{4 \binom{m}{2} \delta_m^2}{m-1} \right)}.$$

Note que

$$\begin{aligned} \frac{1}{m-1} \log \left(1 + \frac{4 \binom{m}{2} \delta_m^2}{m-1} \right) &= \frac{1}{m-1} \log(1 + 2m\delta_m^2) \\ &= \frac{\log m}{m} \left(\frac{m \log(1 + 2m\delta_m^2)}{(\log m)(m-1)} \right) \\ &\stackrel{(i)}{\geq} \frac{\log m}{m} \left(\frac{\log(1 + 2m\delta_m^2)}{\log m} \right) \\ &\stackrel{(ii)}{\geq} \frac{\log m}{m} \left(\frac{\log 2m\delta_m^2}{\log m} \right) \\ &= \frac{\log m}{m} \left(1 + \frac{\log 2\delta_m^2}{\log m} \right), \end{aligned}$$

onde em (i) usamos que $m \geq m - 1$ para todo $m > 0$ e em (ii) usamos que $\log(1 + 2m\delta_m^2) \geq \log 2m\delta_m^2$ para todo $m > 0$. Desse modo, $R^* \geq 1 - \delta_m$ sempre que

$$\mu_m \leq \sqrt{\frac{\log m}{m}} \sqrt{1 + \frac{\log 2\delta_m^2}{\log m}}. \quad (3.21)$$

Por outro lado, pela Proposição 5 temos que $R^* \leq \delta_m$ sempre que

$$\mu_m \geq \sqrt{\frac{2 \log m}{m - 1}} + 2\sqrt{\frac{2 \log(2/\delta_m)}{m - 1}}. \quad (3.22)$$

Suponha que $\delta_m \rightarrow 0$ quando $m \rightarrow \infty$ de modo que $\log \delta_m \ll \log m$. Nesse caso,

$$\sqrt{\frac{\log m}{m}} \sqrt{1 + \frac{\log 2\delta_m^2}{\log m}} \approx \sqrt{\frac{\log m}{m}},$$

quando $m \rightarrow \infty$. Observe que

$$\frac{\log(2/\delta_m)}{m - 1} = \frac{\log 2}{m - 1} - \frac{\log \delta_m}{m - 1},$$

e quando $m \rightarrow \infty$, ambos os termos convergem para zero. Veja que o primeiro termo é essencialmente $1/m$, que converge à zero. Já o segundo termo pode ser visto como

$$\frac{\log \delta_m}{m - 1} = \frac{\log \delta_m}{\log m} \times \frac{\log m}{m - 1},$$

que converge à zero, pela escolha feita para δ_m . Desse modo,

$$\sqrt{\frac{2 \log m}{m - 1}} + 2\sqrt{\frac{2 \log(2/\delta_m)}{m - 1}} \approx \sqrt{\frac{2 \log m}{m - 1}},$$

quando $m \rightarrow \infty$. Portanto, desde que $\delta_m \rightarrow 0$ com $m \rightarrow \infty$ de modo que $\log \delta_m \ll \log m$, temos que para todo $\epsilon > 0$:

- Se $\mu_m \geq (\sqrt{2} + \epsilon)\sqrt{\frac{\log m}{m}}$, então $R^* \rightarrow 0$ quando $m \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.
- Se $\mu_m \leq (1 - \epsilon)\sqrt{\frac{\log m}{m}}$, então $R^* \rightarrow 1$ quando $m \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.

Nesse caso, a ordem de magnitude assintótica para o valor crítico de μ_m é $\sqrt{\frac{\log m}{m}}$.

3.3.4 Árvore geradoras

Considere uma rede com m vértices nos quais cada par de vértices interage. Desejamos testar se existe um subgrafo conectado contendo cada vértice e cujas arestas estão corrompidas, isto é, apresentam comportamento anômalo. Essa situação pode ser modelada considerando a classe das *árvores geradoras*.

Vamos representar as arestas do grafo completo K_m por $1, 2, \dots, n = \binom{m}{2}$ e denotar \mathcal{C}_m o conjunto de todas as árvores geradoras de K_m .

Nesse exemplo, $N_m = m^{m-2}$ (veja Moon (1970)) e $k_m = m - 1$. Pela Proposição 4, temos que $R^* \leq \delta_m$ desde que

$$\begin{aligned}
\mu_m &\geq \sqrt{\frac{8n}{k_m^2} \log(2/\delta_m)} \\
&= \sqrt{\frac{8\binom{m}{2}}{(m-1)^2} \log(2/\delta_m)} \\
&= \sqrt{\frac{4m(m-1)}{(m-1)^2} \log(2/\delta_m)} \\
&= \sqrt{\frac{4m}{m-1} \log(2/\delta_m)} \\
&\geq \sqrt{4 \log(2/\delta_m)}.
\end{aligned} \tag{3.23}$$

Vamos usar a Proposição 8 para obter uma cota inferior do risco de Bayes. Para esse fim, precisamos obter uma cota superior para $\mathbb{E} \left[e^{\mu_m^2 Z} \right]$, onde $Z = |S \cap S'|$ e S, S' são árvores geradoras escolhidas uniformemente e de maneira independente em \mathcal{C}_m . Para isso, precisaremos introduzir algumas notações.

Seja $\Omega = \{0, 1\}^{\{1, \dots, n\}}$, o conjunto de vetores 0/1 indexados por $\{1, \dots, n\}$ e denote por $\omega = (\omega(i) : i \in \{1, \dots, n\})$ um elemento de Ω . Para $\omega, \omega' \in \Omega$, dizemos que $\omega \leq \omega'$ se $\omega(i) \leq \omega'(i)$ para todo $i \in \{1, \dots, n\}$. Uma função $f : \Omega \rightarrow \mathbb{R}$ é dita não decrescente se $f(\omega) \leq f(\omega')$ para todo $\omega \leq \omega'$. No que segue para $F \subseteq \{1, \dots, n\}$ e $\omega \in \Omega$, denotamos $\omega_F = (\omega(i), i \in F)$. Dizemos que f é F-cilíndrica se para quaisquer $\omega, \omega' \in \Omega$ tais que $\omega_F = \omega'_F$, temos $f(\omega) = f(\omega')$.

Note que cada árvore geradora em \mathcal{C}_m pode ser identificada como uma sequência em Ω , onde atribuímos valor 1 para as arestas da árvore geradora e valor 0 para as arestas que não pertencem a árvore geradora. Mais precisamente, cada árvore geradora $s \in \mathcal{C}_m$ é identificada pela sequência $\omega^s = (\omega^s(i), i \in \{1, \dots, n\}) \in \Omega$ definida por $\omega^s(i) = 1$ se e somente se $i \in s$.

Teorema 2. *Seja K_m o grafo completo de m vértices, denote por $1, \dots, n = \binom{m}{2}$ as arestas desse grafo e seja \mathcal{C}_m o conjunto de todas as suas árvores geradoras. Considere funções f e g não decrescentes em Ω tais que f é F -cilíndrica e g é F^c -cilíndrica, onde $F \subseteq \{1, \dots, n\}$. Se S é árvore geradora escolhida uniformemente em \mathcal{C}_m , então*

$$\mathbb{E} \left[f(\omega^S) g(\omega^S) \right] \leq \mathbb{E} \left[f(\omega^S) \right] \mathbb{E} \left[g(\omega^S) \right].$$

Este teorema é atribuído à Feder e Mihail ([Feder and Mihail \(1992\)](#)), cuja prova será omitida nessa dissertação. No que segue, vamos usar o Teorema 2 para obter a cota superior de $\mathbb{E} \left[e^{\mu_m^2 Z} \right]$.

Observe que se S, S' são árvores geradoras escolhidas ao acaso e de maneira independente em \mathcal{C}_m e $Z = |S \cap S'|$, então

$$\begin{aligned} \mathbb{E} \left[e^{\mu_m^2 Z} \right] &= \mathbb{E} \left[\mathbb{E} \left[e^{\mu_m^2 |S \cap S'|} \mid S' \right] \right] \\ &\stackrel{(i)}{=} \mathbb{E} \left[\mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S'} \mathbb{1}_{\{i \in S\}}} \mid S' \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S'} \mathbb{1}_{\{\omega^S(i)=1\}}} \mid S' \right] \right], \end{aligned}$$

onde em (i) usamos que $|S \cap S'| = \sum_{i \in S'} \mathbb{1}_{\{i \in S\}}$. Sejam

$$f(\omega) = e^{\mu_m^2 \sum_{i \in S' \setminus \{j\}} \mathbb{1}_{\{\omega(i)=1\}}} \quad \text{e} \quad g(\omega) = e^{\mu_m^2 \mathbb{1}_{\{\omega(j)=1\}}},$$

para $j \in S'$ fixado e $\omega \in \Omega$. É fácil verificar que f e g são não decrescentes e f é $S' \setminus \{j\}$ -cilíndrica e g é $(S' \setminus \{j\})^c$ -cilíndrica. Assim, pelo Teorema 2 temos que

$$\begin{aligned} \mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S'} \mathbb{1}_{\{\omega^S(i)=1\}}} \mid S' \right] &= \mathbb{E} \left[f(\omega^S) g(\omega^S) \mid S' \right] \\ &\leq \mathbb{E} \left[f(\omega^S) \mid S' \right] \mathbb{E} \left[g(\omega^S) \mid S' \right] \\ &= \mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S' \setminus \{j\}} \mathbb{1}_{\{\omega^S(i)=1\}}} \mid S' \right] \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{\omega^S(j)=1\}}} \mid S' \right]. \end{aligned}$$

Seja $S' \setminus \{j\} = \tilde{S}'$. Tome

$$\tilde{f}(\omega) = e^{\mu_m^2 \sum_{i \in \tilde{S}' \setminus \{\tilde{j}\}} \mathbb{1}_{\{\omega(i)=1\}}} \quad \text{e} \quad \tilde{g}(\omega) = e^{\mu_m^2 \mathbb{1}_{\{\omega(\tilde{j})=1\}}},$$

para algum $\tilde{j} \in S' \setminus \{j\}$ fixado e $\omega \in \Omega$. Pode-se verificar que \tilde{f} e \tilde{g} são não decrescentes e \tilde{f} é $\tilde{S}' \setminus \{\tilde{j}\}$ -cilíndrica e \tilde{g} é $(\tilde{S}' \setminus \{\tilde{j}\})^c$ -cilíndrica. Assim,

aplicando novamente o Teorema 2 temos que

$$\begin{aligned}\mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S' \setminus j} \mathbb{1}_{\{\omega^S(i)=1\}}} | S' \right] &= \mathbb{E} \left[\tilde{f}(\omega^S) \tilde{g}(\omega^S) | S' \right] \\ &\leq \mathbb{E} \left[\tilde{f}(\omega^S) | S' \right] \mathbb{E} \left[\tilde{g}(\omega^S) | S' \right] \\ &= \mathbb{E} \left[e^{\mu_m^2 \sum_{i \in \tilde{S}' \setminus \{\tilde{j}\}} \mathbb{1}_{\{\omega^S(i)=1\}}} | S' \right] \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{\omega^S(\tilde{j})=1\}}} | S' \right].\end{aligned}$$

Aplicando o Teorema 2, $k_m - 2$ vezes temos que

$$\mathbb{E} \left[e^{\mu_m^2 \sum_{i \in S'} \mathbb{1}_{\{\omega^S(i)=1\}}} | S' \right] \leq \prod_{i \in S'} \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{\omega^S(i)=1\}}} | S' \right],$$

e desse modo segue que

$$\begin{aligned}\mathbb{E} \left[e^{\mu_m^2 Z} \right] &\leq \mathbb{E} \left[\prod_{i \in S'} \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{\omega^S(i)=1\}}} | S' \right] \right] \\ &= \mathbb{E} \left[\prod_{i \in S'} \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{i \in S\}}} | S' \right] \right].\end{aligned}$$

Veja que, pela independência de S e S' , sabemos que

$$\mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{i \in S\}}} | i \in S' \right] = \mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{i \in S\}}} \right]. \quad (3.24)$$

Além disso, temos

$$\begin{aligned}\mathbb{E} \left[e^{\mu_m^2 \mathbb{1}_{\{i \in S\}}} \right] &= \mathbb{P}\{i \notin S\} + e^{\mu_m^2} \mathbb{P}\{i \in S\} \\ &= 1 + (e^{\mu_m^2} - 1) \mathbb{P}\{i \in S\}.\end{aligned} \quad (3.25)$$

Para calcular $\mathbb{P}\{i \in S\}$, veja que

$$\mathbb{P}\{i \in S\} = \mathbb{P}\{w^S(i) = 1\} = \frac{|\{\omega^s \in \Omega : s \in \mathcal{C}_m \text{ e } w^s(i) = 1\}|}{N_m}.$$

Note que como o grafo K_m é completo, para quaisquer arestas i e j temos

$$|\{\omega^s \in \Omega : s \in \mathcal{C}_m \text{ e } w^s(i) = 1\}| = |\{\omega^s \in \Omega : s \in \mathcal{C}_m \text{ e } w^s(j) = 1\}|.$$

Denotando $|\{s \in \mathcal{C}_m : w^s(i) = 1\}| = x$, segue que

$$nx = \sum_{i=1}^n \sum_{s \in \mathcal{C}_m} w^s(i) = \sum_{s \in \mathcal{C}_m} \sum_{i=1}^n w^s(i) = N_m(m-1).$$

Logo

$$x = \frac{N_m(m-1)}{n} = \frac{N_m(m-1)}{\binom{m}{2}} = 2m^{m-3}.$$

Portanto, temos que

$$\mathbb{P}\{i \in S\} = \frac{2m^{m-3}}{N_m} = \frac{2m^{m-3}}{m^{m-2}} = \frac{2}{m}. \quad (3.26)$$

Desse modo, veja que

$$\begin{aligned} \mathbb{E} \left[e^{\mu_m^2 Z} \right] &\leq \mathbb{E} \left[\prod_{i \in S'} \mathbb{E} \left[e^{\mu_m^2 1_{\{i \in S\}}} | S' \right] \right] \\ &\stackrel{(ii)}{\leq} \mathbb{E} \left[\prod_{i \in S'} \left(\frac{2(e^{\mu_m^2} - 1)}{m} + 1 \right) \right] \\ &= \left(\frac{2(e^{\mu_m^2} - 1)}{m} + 1 \right)^{m-1} \\ &\stackrel{(iii)}{\leq} \exp \left(\frac{2(m-1)(e^{\mu_m^2} - 1)}{m} \right) \\ &\leq \exp \left(2(e^{\mu_m^2} - 1) \right), \end{aligned}$$

onde em (ii) usamos a inequação (3.24) e substituímos a expressão de $\mathbb{P}\{i \in S\}$ obtida em (3.26) e em (iii) usamos que $1 + x \leq e^x$ para todo $x \in \mathbb{R}$. Assim, pela Proposição 8 temos que $R^* \geq 1 - \delta_m$ quando

$$1 - \frac{1}{2} \sqrt{\exp(2(e^{\mu_m^2} - 1)) - 1} \geq 1 - \delta_m.$$

Isolando μ_m acima temos que $R^* \geq 1 - \delta_m$ desde que

$$\mu_m \leq \sqrt{\log \left(1 + \frac{1}{2} \log(1 + 4\delta_m^2) \right)}. \quad (3.27)$$

Portanto, se $\delta_m \rightarrow 0$ quando $m \rightarrow \infty$ temos que:

- Se $\mu_m \geq \sqrt{4 \log(2/\delta_m)}$, então $R^* \rightarrow 0$ quando $n \rightarrow \infty$ que é equivalente há dizer que é possível detectar anomalias.
- Se $\mu_m \leq \sqrt{\log(1 + \frac{1}{2} \log(1 + 4\delta_m^2))}$, então $R^* \rightarrow 1$ quando $n \rightarrow \infty$ que é equivalente há dizer que é impossível detectar anomalias.

Em particular, se $\liminf \mu_m > 0$, então $R^* \not\rightarrow 1$ quando $n \rightarrow \infty$. Além disso, não obtemos ordem de magnitude assintótica para o valor crítico de μ_m .

3.4 Monotonicidade do risco de Bayes

Intuitivamente, poderíamos esperar que o problema de testar a presença de variáveis anômalas ficaria mais difícil conforme a classe \mathcal{C}_n ficasse maior. Mais precisamente, poderíamos esperar que se $\mathcal{A}_n \subset \mathcal{C}_n$ para duas classes de subconjuntos de $\{1, \dots, n\}$, então $R_{\mathcal{A}_n}^*(\mu_n) \leq R_{\mathcal{C}_n}^*(\mu_n)$ valeria para todo μ_n , onde $R_{\mathcal{A}_n}^*(\mu_n)$ é o risco de Bayes associado a classe \mathcal{A}_n . Vamos mostrar que essa condição, em geral não vale. Vamos mostrar que existem classes $\mathcal{A}_n \subset \mathcal{C}_n$ para as quais $R_{\mathcal{A}_n}^*(\mu_n)$ é arbitrariamente perto de 1 e $R_{\mathcal{C}_n}^*(\mu_n)$ é arbitrariamente perto de 0 para o mesmo valor de μ_n . No entanto, vamos mostrar que a monotonicidade vale para classes \mathcal{C}_n suficientemente simétricas.

Dizemos que a classe \mathcal{C}_n é simétrica se para o teste ótimo

$$f_{\mathcal{C}_n}^*(x) = \mathbb{1}_{\{\frac{1}{N_n} \sum_{S' \in \mathcal{C}_n} \exp(\mu_n \sum_{i \in S'} x_i) \geq \exp(k_n \mu_n^2 / 2)\}},$$

o valor de $\mathbb{P}_S\{f_{\mathcal{C}_n}^*(X) = 0\}$ é o mesmo para todo $S \in \mathcal{C}_n$. Note que alguns dos exemplos discutidos na Seção 3.3 satisfazem a hipótese de simetria, como as classes dos k_n -conjuntos e estrelas.

Teorema 3 (Addario-Berry et al. (2010)). *Seja \mathcal{C}_n uma classe simétrica de subconjuntos de $\{1, \dots, n\}$. Se \mathcal{A}_n é uma subclasse arbitrária de \mathcal{C}_n , então $R_{\mathcal{A}_n}^*(\mu_n) \leq R_{\mathcal{C}_n}^*(\mu_n)$ para todo $\mu_n > 0$.*

Demonstração. Fixe $\mu_n > 0$ e omita ele na notação. Vamos denotar aqui o risco do teste do pior caso possível como sendo

$$\overline{R_{\mathcal{C}_n}}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S\{f(X) = 0\}.$$

Lembre que o risco médio é dado por

$$R_{\mathcal{C}_n}(f) = \mathbb{P}_0\{f(X) = 1\} + \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f(X) = 0\}.$$

Vamos verificar que o risco do pior caso possível é monótono no sentido de que se $\mathcal{A}_n \subset \mathcal{C}_n$ então para cada f , $\overline{R_{\mathcal{A}_n}}(f) \leq \overline{R_{\mathcal{C}_n}}(f)$. Como $\mathcal{A}_n \subset \mathcal{C}_n$, logo

$$\max_{S \in \mathcal{A}_n} \mathbb{P}_S\{f(X) = 0\} \leq \max_{S' \in \mathcal{C}_n} \mathbb{P}_{S'}\{f(X) = 0\},$$

e segue que para cada teste f temos que $\overline{R_{\mathcal{A}_n}}(f) \leq \overline{R_{\mathcal{C}_n}}(f)$.

Sejam $f_{\mathcal{C}_n}^*$ e $f_{\mathcal{A}_n}^*$ os testes ótimos com respeito ao risco do pior caso possível e o risco médio, respectivamente. Como \mathcal{C}_n é simétrica, logo $\mathbb{P}_S\{f_{\mathcal{C}_n}^*(X) = 0\}$

é o mesmo para todo $S \in \mathcal{C}_n$ e então $\overline{R_{\mathcal{C}_n}}(f_{\mathcal{C}_n}^*) = R_{\mathcal{C}_n}(f_{\mathcal{C}_n}^*)$. Já que $R_{\mathcal{C}_n}(f) \leq \overline{R_{\mathcal{C}_n}}(f)$ para todo f , temos que

$$\overline{R_{\mathcal{C}_n}}(\overline{f_{\mathcal{C}_n}}) \leq \overline{R_{\mathcal{C}_n}}(f_{\mathcal{C}_n}^*) = R_{\mathcal{C}_n}(f_{\mathcal{C}_n}^*) \leq R_{\mathcal{C}_n}(\overline{f_{\mathcal{C}_n}}) \leq \overline{R_{\mathcal{C}_n}}(\overline{f_{\mathcal{C}_n}}).$$

Logo, todas as desigualdades são igualdades e em particular, $\overline{f_{\mathcal{C}_n}} = f_{\mathcal{C}_n}^*$. Como \mathcal{A}_n é uma subclasse arbitrária de \mathcal{C}_n , logo

$$R_{\mathcal{C}_n}^* = R_{\mathcal{C}_n}(f_{\mathcal{C}_n}^*) = \overline{R_{\mathcal{C}_n}}(f_{\mathcal{C}_n}^*) \geq \overline{R_{\mathcal{A}_n}}(\overline{f_{\mathcal{C}_n}}) \geq R_{\mathcal{A}_n}(\overline{f_{\mathcal{C}_n}}) \geq R_{\mathcal{A}_n}(f_{\mathcal{A}_n}^*) = R_{\mathcal{A}_n}^*.$$

□

Quando a hipótese de simetria não é satisfeita, podemos escolher classes $\mathcal{A}_n \subset \mathcal{C}_n$ para as quais $R_{\mathcal{A}_n}^*(\mu_n)$ é arbitrariamente perto de 1 e $R_{\mathcal{C}_n}^*(\mu_n)$ é arbitrariamente perto de 0 para o mesmo valor de μ_n . Nesse caso, o risco de Bayes não é monótono.

Teorema 4 (Addario-Berry et al. (2010)). *Para cada $\epsilon \in (0, 1)$, existem n , μ_n e classes $\mathcal{A}_n \subset \mathcal{C}_n \subset \{1, \dots, n\}$ tais que $R_{\mathcal{A}_n}^*(\mu_n) \geq 1 - \epsilon$ e $R_{\mathcal{C}_n}^*(\mu_n) \leq 2\epsilon$.*

Demonstração. Para cada classe \mathcal{L}_n , denote $\phi_{\mathcal{L}_n}(x) = \frac{1}{N_n} \sum_{S \in \mathcal{L}_n} \phi_S(x)$, onde ϕ_S é definido em (2.6). Lembre que da Proposição 2

$$R_{\mathcal{L}_n}^*(\mu_n) = 1 - \frac{1}{2} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{L}_n}(x)| dx,$$

onde ϕ_0 é definido em (2.5).

Dado $\epsilon \in (0, 1)$, fixe $k_n = k_n(\epsilon)$ grande o suficiente para que $k_n + 1 \geq \frac{1}{\epsilon}$ e

$$\sqrt{\frac{\log(4(k_n + 1)\epsilon^2)}{k_n + 1}} \geq \sqrt{\frac{8}{k_n} \log\left(\frac{2}{\epsilon}\right)},$$

e seja $n = n(\epsilon) = (k_n + 1)^2$. Seja \mathcal{A}_n a escolha de $k_n + 1$ conjuntos disjuntos de $\{1, \dots, n\}$ e cada um com tamanho $k_n + 1$. Seja \mathcal{B}_n todos os conjuntos da forma $\{1, \dots, k_n, i\}$ onde i varia de $k_n + 1$ até n e assumamos que \mathcal{A}_n é escolhido de tal forma que $\mathcal{A}_n \cap \mathcal{B}_n = \emptyset$. Agora, faça $\mathcal{C}_n = \mathcal{A}_n \cup \mathcal{B}_n$. Do exemplo dos conjuntos disjuntos, encontrado em 3.3.1, temos que $R_{\mathcal{A}_n}^*(\mu_n) \geq 1 - \epsilon$ vale quando

$$\mu_n \leq \sqrt{\frac{\log(4N_n\epsilon^2)}{k_n + 1}} = \sqrt{\frac{\log(4(k_n + 1)\epsilon^2)}{k_n + 1}}.$$

Então, tomando $\mu_n = \sqrt{\frac{\log(4(k_n + 1)\epsilon^2)}{k_n + 1}}$ vale que $R_{\mathcal{A}_n}^*(\mu_n) \geq 1 - \epsilon$. Vamos obter uma cota superior para $R_{\mathcal{B}_n}^*(\mu_n)$ usando o teste baseado na soma,

$$f_{TS}(x) = \mathbb{1}_{\left\{\sum_{i=1}^{k_n} x_i > \mu_n k_n / 2\right\}}.$$

De maneira análoga a demonstração da Proposição 4, temos que $R(f_{TS}) \leq \epsilon$ quando $\mu_n \geq \sqrt{\frac{8}{k_n} \log\left(\frac{2}{\epsilon}\right)}$. Pelas escolhas feitas por μ_n e k_n , temos que $\mu_n \geq \sqrt{\frac{8}{k_n} \log\left(\frac{2}{\epsilon}\right)}$. Assim, $R_{\mathcal{B}_n}^*(\mu_n) \leq R(f_{TS}) \leq \epsilon$. Lembre que da Proposição 2 $R_{\mathcal{A}_n}^*(\mu_n) = 1 - \frac{1}{2} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{A}_n}(x)| dx$. Logo, pela escolha feita para μ_n e k_n temos que

$$\int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{A}_n}(x)| dx = 2 - 2R_{\mathcal{A}_n}^*(\mu_n) \leq 2 - 2(1 - \epsilon) = 2\epsilon.$$

Seja $M = |\mathcal{B}_n| = (k_n + 1)^2 - k_n$. Desse modo, $N_n = |\mathcal{C}_n| = |\mathcal{A}_n| + |\mathcal{B}_n| = k_n + 1 + M = (k_n + 1)^2 + 1$. Note que

$$\begin{aligned} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{C}_n}(x)| dx &= \int_{\mathbb{R}^n} \left| \phi_0(x) - \frac{(k_n + 1)\phi_{\mathcal{A}_n}(x) + M\phi_{\mathcal{B}_n}(x)}{N_n} \right| dx \\ &= \int_{\mathbb{R}^n} \left| \frac{N_n\phi_0(x) - (k_n + 1)\phi_{\mathcal{A}_n}(x) - M\phi_{\mathcal{B}_n}(x)}{N_n} \right| dx \\ &= \int_{\mathbb{R}^n} \left| \frac{(k_n + 1)(\phi_0(x) - \phi_{\mathcal{A}_n}(x)) + M(\phi_0(x) - \phi_{\mathcal{B}_n}(x))}{N_n} \right| dx \\ &\stackrel{(i)}{\geq} \int_{\mathbb{R}^n} \left| \frac{M(\phi_0(x) - \phi_{\mathcal{B}_n}(x))}{N_n} \right| dx \\ &\quad - \int_{\mathbb{R}^n} \left| \frac{(k_n + 1)(\phi_0(x) - \phi_{\mathcal{A}_n}(x))}{N_n} \right| dx \\ &= \frac{M}{N_n} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{B}_n}(x)| dx \\ &\quad - \frac{k_n + 1}{N_n} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{A}_n}(x)| dx \\ &\stackrel{(ii)}{\geq} \frac{M}{N_n} (2 - 2R_{\mathcal{B}_n}^*(\mu_n)) - \frac{k_n + 1}{N_n} 2\epsilon \\ &\geq \frac{M}{N_n} (2 - 2\epsilon) - \frac{k_n + 1}{N_n} 2\epsilon \\ &= \frac{2M - 2\epsilon N_n}{N_n} \\ &= \frac{2M}{N_n} - 2\epsilon \\ &\stackrel{(iii)}{=} 2 - 2\frac{k_n + 1}{N_n} - 2\epsilon \\ &\stackrel{(iv)}{\geq} 2 - 4\epsilon, \end{aligned}$$

onde em (i) usamos a desigualdade triangular, em (ii) usamos a desigualdade $R_{\mathcal{B}_n}^*(\mu_n) \leq \epsilon$, em (iii) usamos que $M = N_n - (k_n + 1)$ e em (iv) usamos que

$$\frac{k_n + 1}{N_n} = \frac{k_n + 1}{(k_n + 1)^2 + 1} \leq \frac{1}{k_n + 1} \leq \epsilon.$$

Portanto, usando mais uma vez a Proposição 2 e a cota inferior acima obtida temos que

$$R_{\mathcal{C}_n}^*(\mu_n) = 1 - \frac{1}{2} \int_{\mathbb{R}^n} |\phi_0(x) - \phi_{\mathcal{C}_n}(x)| dx \leq 1 - \frac{1}{2}(2 - 4\epsilon) = 2\epsilon,$$

concluindo a prova. \square

3.5 Cota inferior baseada em subclasses aleatórias

Nesta seção vamos obter uma outra cota inferior para o risco médio $R^* = R_{\mathcal{C}_n}^*$. Vamos tratar \mathcal{C}_n como o espaço métrico equipado com a distância canônica

$$d(S, T) = \sqrt{\mathbb{E}_0 \left[(X_S - X_T)^2 \right]}, \quad (3.28)$$

onde $S, T \in \mathcal{C}_n$ e $X_S = \sum_{i \in S} X_i$.

Para $M_n \leq N_n$ inteiro, definimos o parâmetro $t_{\mathcal{C}_n}(M_n) > 0$ da classe \mathcal{C}_n da seguinte forma. Denote $\mathcal{A}_n \subset \mathcal{C}_n$ o conjunto obtido a partir de M_n extrações sem reposição de \mathcal{C}_n e defina a variável aleatória τ como a menor distância entre elementos de \mathcal{A}_n , isto é,

$$\tau = \min_{S, T \in \mathcal{A}_n, S \neq T} d(S, T).$$

Seja $t_{\mathcal{C}_n}(M_n)$ a mediana de τ .

Teorema 5 (Addario-Berry et al. (2010)). *Seja $M_n \leq N_n$ inteiro. Então para toda classe \mathcal{C}_n ,*

$$R_{\mathcal{C}_n}^* \geq \delta^*/2,$$

sempre que,

$$\mu_n \leq \min \left(\sqrt{\frac{\log((\delta^*)^2 M_n / 4)}{k_n}}, \sqrt{\frac{8 \log \left(\frac{4 - \delta^*}{8 \sqrt{2\delta^* - (\delta^*)^2}} \right)}{k_n - t_{\mathcal{C}_n}(M_n)^2 / 2}} \right),$$

onde $\delta^* > 0$ é o maior valor de $\delta \in (0, 1)$ tal que $\frac{8}{4 - \delta} \sqrt{2\delta - \delta^2} < 1$.

Para interpretar o teorema acima, note que

$$k_n - \tau^2/2 = \max_{S, T \in \mathcal{A}_n, S \neq T} |S \cap T|,$$

é a maior sobreposição entre dois elementos de \mathcal{A}_n . Assim como na Proposição 8, a distribuição da sobreposição entre dois elementos de \mathcal{C}_n é uma quantidade importante para estabelecer cotas inferiores para o risco de Bayes. Lembre que na Proposição 8 é a função geradora de momentos da sobreposição entre dois elementos aleatórios que determinam cotas inferiores para o risco de Bayes, aqui é a mediana da maior sobreposição entre muitos elementos aleatórios que faz esse papel. Observe que o Teorema 5 fornece apenas condições suficientes para que o risco de Bayes não convirja para zero. Em particular, pela definição de δ^* temos que $\delta^* \leq 1/2$ de modo que não é possível fazer o risco de Bayes arbitrariamente próximo de 1.

Vamos considerar a classe \mathcal{C}_m das estrelas apresentado na Seção 3.3. Nesse caso, lembre que $n = \binom{m}{2}$, $N_n = m$ e $k_n = m - 1$. Tome $M_n = m$. Veja que $k_n - \tau^2/2 = \max_{S, T \in \mathcal{A}_n, S \neq T} |S \cap T| = 1$, pois para $S \neq T$ com $S, T \in \mathcal{A}_n$ vale que $|S \cap T| = 1$. Desse modo, $\tau = \sqrt{2(k_n - 1)}$. Logo τ é constante e assim, $t_{\mathcal{C}_m}(M_n) = \sqrt{2(k_n - 1)}$. Pelo Teorema 5 temos que $R_{\mathcal{C}_m}^* \geq \delta^*/2$ sempre que

$$\begin{aligned} \mu_m &\leq \min \left(\sqrt{\frac{\log((\delta^*)^2 m/4)}{m-1}}, \sqrt{8 \log \left(\frac{4 - \delta^*}{8\sqrt{2\delta^* - (\delta^*)^2}} \right)} \right) \\ &\approx \sqrt{\frac{\log((\delta^*)^2 m/4)}{m-1}}, \end{aligned}$$

onde δ^* o maior valor de $\delta \in (0, 1)$ tal que $\frac{8}{4-\delta} \sqrt{2\delta - \delta^2} < 1$. Portanto, se $\mu_m \leq \sqrt{\frac{\log((\delta^*)^2 m/4)}{m-1}}$, então $\liminf R^* > 0$. Na verdade, pelos resultados apresentados no Exemplo 3.3.3, sabemos que $R^* \rightarrow 1$ quando $m \rightarrow \infty$ pois o valor crítico de μ_m é da ordem de $\sqrt{(\log m)/m}$.

Para mostrar o Teorema 5, vamos precisar dos seguintes resultados.

Lema 6 (Desigualdade de Slepian, forma funcional). *Considere dois vetores aleatórios Gaussianos X e Y em \mathbb{R}^n com média zero. Assuma que para todo $i, j = 1, \dots, n$, temos que*

$$\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2] \quad e \quad \mathbb{E}[X_i X_j] \leq \mathbb{E}[Y_i Y_j].$$

Considere uma função duas vezes diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0 \quad \text{para todo } i \neq j.$$

Então,

$$\mathbb{E} [f(X)] \geq \mathbb{E} [f(Y)].$$

A prova do Lema 6 é feita no Apêndice B.

Lema 7. *Seja $M_n \leq N_n$ inteiro fixo e escolha M_n conjuntos de \mathcal{C}_n de maneira uniforme e sem reposição. Denote por \mathcal{A}_n a subclasse de \mathcal{C}_n obtida dessa maneira e $\hat{\mathbb{E}}$ a esperança com respeito a escolha aleatória \mathcal{A}_n . Seja $V_S = e^{\mu_n X_S - k_n \mu_n^2 / 2}$, onde $X_S = \sum_{i \in S} X_i$. Então,*

$$\begin{aligned} \mathbb{E}_0 \left[\hat{\mathbb{E}} \left[\left(\frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \right] \\ \leq \mathbb{E}_0 \left[\frac{1}{M_n} \frac{1}{N_n} \sum_{T \in \mathcal{C}_n} \left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right]. \end{aligned} \quad (3.29)$$

Demonstração. Seja $Y_S = \mathbb{1}_{\{S \in \mathcal{A}_n\}} V_S$. Logo $Y_S \in \{0, V_S\}$ e $\hat{\mathbb{P}}\{V_S = Y_S\} = \frac{\binom{N_n-1}{M_n-1}}{\binom{N_n}{M_n}} = \frac{M_n}{N_n}$. Veja que

$$\begin{aligned} \mathbb{E}_0 \left[\hat{\mathbb{E}} \left[\left(\frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \right] &= \mathbb{E}_0 \left[\hat{\mathbb{E}} \left[\left(\frac{1}{M_n} \sum_{S \in \mathcal{C}_n} \mathbb{1}_{\{S \in \mathcal{A}_n\}} V_S \right. \right. \right. \\ &\quad \left. \left. \left. - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \right] \\ &\stackrel{(i)}{=} \mathbb{E}_0 \left[\widehat{Var} \left(\frac{1}{M_n} \sum_{S \in \mathcal{C}_n} \mathbb{1}_{\{S \in \mathcal{A}_n\}} V_S \right) \right], \end{aligned} \quad (3.30)$$

onde em (i) usamos que $\hat{\mathbb{E}} \left[\frac{1}{M_n} \sum_{S \in \mathcal{C}_n} \mathbb{1}_{\{S \in \mathcal{A}_n\}} V_S \right] = \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S$. Note que

$$\widehat{Var} \left(\frac{1}{M_n} \sum_{S \in \mathcal{C}_n} Y_S \right) = \frac{1}{M_n^2} \sum_{S \in \mathcal{C}_n} \widehat{Var} (Y_S) + \frac{1}{M_n^2} \sum_{S \neq T} \widehat{Cov} (Y_S, Y_T). \quad (3.31)$$

Agora, observe que

$$\widehat{Var}(Y_S) = V_S^2 \frac{M_n}{N_n} \left(1 - \frac{M_n}{N_n}\right), \quad (3.32)$$

e

$$\begin{aligned} \widehat{Cov}(Y_S, Y_T) &= V_S V_T \widehat{Cov}(\mathbb{1}_{\{S \in \mathcal{A}_n\}}, \mathbb{1}_{\{T \in \mathcal{A}_n\}}) \\ &= V_S V_T \left(\widehat{\mathbb{P}}\{S \in \mathcal{A}_n, T \in \mathcal{A}_n\} - \left(\frac{M_n}{N_n}\right)^2 \right) \\ &= V_S V_T \left(\frac{M_n}{N_n} \frac{M_n - 1}{N_n - 1} - \left(\frac{M_n}{N_n}\right)^2 \right). \end{aligned} \quad (3.33)$$

Após substituir (3.32) e (3.33) na igualdade (3.31) e fazer alguns cálculos, concluímos que

$$\widehat{Var}\left(\frac{1}{M_n} \sum_{S \in \mathcal{C}_n} Y_S\right) = \frac{1}{M_n N_n} \sum_{S \in \mathcal{C}_n} V_S^2 + \frac{M_n - 1}{N_n - 1} \frac{1}{M_n N_n} \sum_{S \neq T} V_S V_T - \frac{1}{N_n^2} \sum_{S, T} V_S V_T. \quad (3.34)$$

Por outro lado,

$$\begin{aligned} \frac{1}{M_n} \frac{1}{N_n} \sum_{T \in \mathcal{C}_n} \left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 &= \frac{1}{M_n N_n} \sum_{T \in \mathcal{C}_n} V_T^2 - \frac{2}{M_n} \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \\ &\quad + \frac{1}{M_n} \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \\ &= \frac{1}{M_n N_n} \sum_{T \in \mathcal{C}_n} V_T^2 - \frac{1}{M_n} \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2. \end{aligned}$$

Logo, temos que

$$\mathbb{E}_0 \left[\frac{1}{M_n} \frac{1}{N_n} \sum_{T \in \mathcal{C}_n} \left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] = \mathbb{E}_0 \left[\frac{1}{M_n N_n} \sum_{T \in \mathcal{C}_n} V_T^2 - \frac{1}{M_n} \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right].$$

Substituindo (3.34) em (3.30) e usando a igualdade acima, vemos que para verificar (3.31) é preciso demonstrar a desigualdade

$$\mathbb{E}_0 \left[\frac{M_n - 1}{N_n - 1} \frac{1}{M_n N_n} \sum_{S \neq T} V_S V_T - \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \leq \mathbb{E}_0 \left[-\frac{1}{M_n} \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right].$$

Após isolar os termos com a soma mista e a soma dos quadrados de V_S e ver que

$$\left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 = \frac{1}{N_n^2} \sum_{S \in \mathcal{C}_n} V_S^2 + \frac{1}{N_n^2} \sum_{S \neq T} V_S V_T,$$

observamos que o que queremos mostrar é equivalente à

$$\mathbb{E}_0 \left[\frac{1}{N_n(N_n - 1)} \sum_{S \neq T} V_S V_T \right] \leq \mathbb{E}_0 \left[\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S^2 \right].$$

Note que para todo $T, S \in \mathcal{C}_n$,

$$\begin{aligned} \mathbb{E}_0 [V_S V_T] &= \mathbb{E}_0 \left[e^{\mu_n(X_S + X_T)} \right] e^{-k_n \mu_n^2} \\ &= e^{\mu_n^2 k_n - \mu_n^2 |T \cap S| + 2\mu_n^2 |T \cap S|} e^{-k_n \mu_n^2} \\ &= e^{\mu_n^2 |T \cap S|}. \end{aligned}$$

Em particular, para $T = S$ temos que $\mathbb{E}_0 [V_S^2] = e^{\mu_n^2 k_n}$. Desse modo, basta mostrar que

$$\frac{1}{N_n(N_n - 1)} \sum_{S \neq T} e^{\mu_n^2 |T \cap S|} \leq e^{\mu_n^2 k_n}.$$

Como $|T \cap S| \leq k_n$, temos que

$$\begin{aligned} \frac{1}{N_n(N_n - 1)} \sum_{S \neq T} e^{\mu_n^2 |T \cap S|} &\leq \frac{1}{N_n(N_n - 1)} \sum_{S \neq T} e^{\mu_n^2 k_n} \\ &= \frac{1}{N_n(N_n - 1)} N_n(N_n - 1) e^{\mu_n^2 k_n} \\ &= e^{\mu_n^2 k_n}, \end{aligned}$$

concluindo a prova. □

Demonstração do Teorema 5. Seja $M_n \leq N_n$ inteiro fixo e escolha M_n conjuntos de \mathcal{C}_n de maneira uniforme e sem reposição. Denote por \mathcal{A}_n a subclasse de \mathcal{C}_n obtida dessa maneira. Denote a razão de verossimilhança associada a essa classe por

$$L_{\mathcal{A}_n}(X) = \frac{(1/M_n) \sum_{S \in \mathcal{A}_n} \phi_S(X)}{\phi_0(x)} = \frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S,$$

onde $V_S = e^{\mu_n X_S - k_n \mu_n^2 / 2}$, ϕ_0 é definido em (2.5) e ϕ_S é definido em (2.6). Assim, podemos dar cota inferior para o risco de Bayes da classe \mathcal{C}_n notando que pela Proposição 2 vale que

$$\begin{aligned} R_{\mathcal{C}_n}^*(\mu_n) - R_{\mathcal{A}_n}^*(\mu_n) &= \frac{1}{2} \left(\mathbb{E}_0 [|L_{\mathcal{A}_n}(X) - 1|] - \mathbb{E}_0 [|L_{\mathcal{C}_n}(X) - 1|] \right) \\ &\stackrel{(i)}{\geq} -\frac{1}{2} \mathbb{E}_0 [|L_{\mathcal{A}_n}(X) - L_{\mathcal{C}_n}(X)|], \end{aligned}$$

onde em (i) usamos a desigualdade triangular. Denotando por $\hat{\mathbb{E}}$, a esperança com respeito a escolha aleatória \mathcal{A}_n , temos que

$$\begin{aligned} R_{\mathcal{C}_n}^*(\mu_n) &\stackrel{(ii)}{\geq} \hat{\mathbb{E}} [R_{\mathcal{A}_n}^*(\mu_n)] - \frac{1}{2} \mathbb{E}_0 \left[\hat{\mathbb{E}} \left[\left| \frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right| \right] \right] \\ &\stackrel{(iii)}{\geq} \hat{\mathbb{E}} [R_{\mathcal{A}_n}^*(\mu_n)] - \frac{1}{2} \sqrt{ \mathbb{E}_0 \left[\hat{\mathbb{E}} \left[\left(\frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \right] } \\ &\stackrel{(iv)}{\geq} \hat{\mathbb{E}} [R_{\mathcal{A}_n}^*(\mu_n)] - \frac{1}{2} \sqrt{ \mathbb{E}_0 \left[\frac{1}{M_n} \frac{1}{N_n} \sum_{T \in \mathcal{C}_n} \left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] } \\ &= \hat{\mathbb{E}} [R_{\mathcal{A}_n}^*(\mu_n)] - \frac{1}{2\sqrt{M_n}} \sqrt{ \frac{1}{N_n} \sum_{T \in \mathcal{C}_n} \mathbb{E}_0 \left[\left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] } \end{aligned}$$

onde em (ii) usamos o Teorema de Fubini, em (iii) usamos a desigualdade de Cauchy-Schwarz e a desigualdade de Jensen, e em (iv) usamos o Lema 7.

Podemos limitar o termo da direita notando que

$$\begin{aligned}
\mathbb{E}_0 \left[\left(V_T - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] &\stackrel{(v)}{\leq} 2\mathbb{E}_0 \left[(V_T - 1)^2 \right] + 2\mathbb{E}_0 \left[\left(1 - \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} V_S \right)^2 \right] \\
&\stackrel{(vi)}{=} 2\mathbb{E}_0 \left[(V_T - 1)^2 \right] + 2\mathbb{E}_0 \left[\left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} (V_S - 1) \right)^2 \right] \\
&= 2\mathbb{E}_0 \left[(V_T - 1)^2 \right] + \frac{2}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{E}_0 \left[(V_S - 1)^2 \right] \\
&\stackrel{(vii)}{=} 2\text{Var}(V_T) + \frac{2}{N_n} \sum_{S \in \mathcal{C}_n} \text{Var}(V_S) \\
&= 4\text{Var}(V_T) \\
&\stackrel{(viii)}{=} 4 \left(e^{\mu_n^2 k_n} - 1 \right),
\end{aligned}$$

onde em (v) usamos a cota $(a - b)^2 \leq 2(a^2 + b^2)$, para $a, b > 0$, em (vi) usamos que $N_n = |\mathcal{C}_n|$, em (vii) usamos que $\mathbb{E}_0[V_T] = 1$ e em (viii) usamos a função geradora de momentos de V_T e que $V_T = e^{\mu_n X_T - k_n \mu_n^2 / 2}$. Em conclusão, temos que

$$R_{\mathcal{C}_n}^*(\mu_n) \geq \hat{\mathbb{E}} \left[R_{\mathcal{A}_n}^*(\mu_n) \right] - \sqrt{\frac{e^{\mu_n^2 k_n} - 1}{M_n}} \stackrel{(ix)}{\geq} \hat{\mathbb{E}} \left[R_{\mathcal{A}_n}^*(\mu_n) \right] - \frac{\delta_n}{2},$$

onde em (ix) usamos a hipótese $\mu_n \leq \sqrt{(1/k_n) \log(\delta_n^2 M_n / 4)}$. Logo, é suficiente mostrar que $\hat{\mathbb{E}} \left[R_{\mathcal{A}_n}^*(\mu_n) \right] \geq \delta$.

Vamos limitar o risco de Bayes associado a \mathcal{A}_n em termos da afinidade de Bhattacharyya,

$$\rho_{\mathcal{A}_n}(\mu_n) = \frac{1}{2} \mathbb{E}_0 \left[\sqrt{L_{\mathcal{A}_n}(X)} \right] = \frac{1}{2} \mathbb{E}_0 \left[\sqrt{\frac{1}{M_n} \sum_{S \in \mathcal{A}_n} V_S} \right].$$

Lembre que $R_{\mathcal{A}_n}^*(\mu_n) \geq 1 - \sqrt{1 - 4\rho_{\mathcal{A}_n}(\mu_n)^2}$ e usando o fato da função $x \mapsto \sqrt{1 - 4x^2}$ ser côncava, temos pela desigualdade de Jensen que

$$\hat{\mathbb{E}} \left[R_{\mathcal{A}_n}^*(\mu_n) \right] \geq 1 - \sqrt{1 - 4 \left(\hat{\mathbb{E}} \left[\rho_{\mathcal{A}_n}(\mu_n) \right] \right)^2}.$$

Desse modo, temos que demonstrar que

$$\hat{\mathbb{E}} [\rho_{\mathcal{A}_n}(\mu_n)] \geq \frac{\sqrt{2\delta - \delta^2}}{2}.$$

No argumento abaixo, fixamos uma classe aleatória \mathcal{A}_n , renomeamos seus elementos de modo que $\mathcal{A}_n = \{1, 2, \dots, |\mathcal{A}_n|\}$ e limitamos $\rho_{\mathcal{A}_n}(\mu_n)$ por baixo. Denote a mínima distância entre quaisquer dois elementos de \mathcal{A}_n por τ . Para limitar $\rho_{\mathcal{A}_n}(\mu_n)$, vamos aplicar o Lema 6 (Slepian) com a função

$$F(\mathbf{x}) = \sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{i=1}^{|\mathcal{A}_n|} e^{\mu_n x_i - k_n \mu_n^2 / 2}},$$

onde $\mathbf{x} = (x_1, \dots, x_{|\mathcal{A}_n|})$. Um simples cálculo mostra que

$$\frac{\partial^2 F(\mathbf{x})}{\partial x_i \partial x_j} = -\frac{1}{4} \frac{\mu_n^2 e^{\mu_n(x_i + x_j) - k_n \mu_n^2 / 2}}{\sqrt{F(\mathbf{x})^3}} \leq 0,$$

para todo $i \neq j$ e para todo $\mathbf{x} = (x_1, \dots, x_{|\mathcal{A}_n|})$. Portanto, podemos aplicar o Lema 6.

Para esse fim, vamos introduzir dois vetores aleatórios $\xi = (\xi_S)_{S \in \mathcal{A}_n}$ e $\zeta = (\zeta_S)_{S \in \mathcal{A}_n}$ em $\mathbb{R}^{|\mathcal{A}_n|}$ Gaussianos com média zero. Definimos $\xi_S = X_S$ para todo $S \in \mathcal{A}_n$. Então sob \mathbb{P}_0 , ξ_S segue distribuição normal com média 0 e variância k_n , e além disso, $\mathbb{E}_0 [F(\xi)] = 2\rho_{\mathcal{A}_n}(\mu_n)$. Para definir o vetor aleatório ζ , introduzimos $M_n + 1$ variáveis aleatórias independentes com distribuição normal padrão da seguinte maneira: uma variável G_S para cada $S \in \mathcal{A}_n$ e uma variável extra G_0 . Definimos para cada $S \in \mathcal{A}_n$,

$$\zeta_S = G_S \frac{\tau}{\sqrt{2}} + G_0 \sqrt{k_n - \frac{\tau^2}{2}}.$$

Logo,

$$\mathbb{E}_0 [\zeta_S \zeta_T] = \begin{cases} k_n, & \text{se } S = T, \\ k_n - \frac{\tau^2}{2}, & \text{se } S \neq T. \end{cases}$$

Veja que $\mathbb{E}_0 [\xi_S \xi_T] = \mathbb{E}_0 [X_S X_T] = |S \cap T|$. Por outro lado,

$$d^2(S, T) = \mathbb{E}_0 [(X_S - X_T)^2] = 2k_n^2 - 2\mathbb{E}_0 [X_S X_T] = 2k_n^2 - 2|S \cap T|.$$

Assim, $\mathbb{E}_0 [\xi_S^2] = k_n$ e quando $S \neq T$,

$$\mathbb{E}_0 [\xi_S \xi_T] = |S \cap T| = k_n - \frac{d(S, T)^2}{2} \stackrel{(ix)}{\leq} k_n - \frac{\tau^2}{2} = \mathbb{E}_0 [\zeta_S \zeta_T],$$

onde em (ix) usamos a definição de τ .

Portanto, pelo Lema 6, temos que $2\rho_{\mathcal{A}_n}(\mu_n) = \mathbb{E}_0 [F(\xi)] \geq \mathbb{E}_0 [F(\zeta)]$. No entanto,

$$\begin{aligned}
\mathbb{E}_0 [F(\zeta)] &= \mathbb{E}_0 \left[\sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n \zeta_S - k_n \mu_n^2 / 2}} \right] \\
&= \mathbb{E}_0 \left[\sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n G_S \frac{\tau}{2} + \mu_n G_0 \sqrt{k_n - \frac{\tau^2}{2} - \frac{k_n \mu_n^2}{2} + \frac{\mu_n^2 \tau^2}{4} - \frac{\mu_n^2 \tau^2}{4}}} \right] \\
&= \mathbb{E}_0 \left[\sqrt{e^{\mu_n G_0 \sqrt{k_n - \frac{\tau^2}{2} - (k_n - \frac{\tau^2}{2}) \mu_n^2 / 2}} \frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n \tau \frac{G_S}{\sqrt{2}} - \frac{\tau^2 \mu_n^2}{4}}} \right] \\
&\stackrel{(x)}{=} \mathbb{E}_0 \left[e^{\mu_n G_0 \sqrt{k_n - \frac{\tau^2}{2} / 2 - (k_n - \frac{\tau^2}{2}) \mu_n^2 / 4}} \right] \mathbb{E} \left[\sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n \tau \frac{G_S}{\sqrt{2}} - \frac{\tau^2 \mu_n^2}{4}}} \right] \\
&\stackrel{(xi)}{=} e^{-\mu_n^2 (k_n - \tau^2 / 2) / 8} \mathbb{E}_0 \left[\sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n \tau \frac{G_S}{\sqrt{2}} - \frac{\tau^2 \mu_n^2}{4}}} \right],
\end{aligned}$$

onde em (x) usamos que G_0 e G_S são independentes para todo $S \in \mathcal{A}_n$, e em (xi) usamos a função geradora de momentos de G_0 . Veja que a última expressão é a afinidade de Bhattacharyya correspondente a classe de conjuntos disjuntos, todos de tamanho $k_n = \tau^2 / 2$ e de cardinalidade $|\mathcal{A}_n| = M_n$. Logo, pela Proposição 3 temos que para $\delta \in (0, 1)$

$$\mathbb{E}_0 \left[\sqrt{\frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} e^{\mu_n \tau \frac{G_S}{\sqrt{2}} - \frac{\tau^2 \mu_n^2}{4}}} \right] \geq R_{\mathcal{A}_n}^* \stackrel{(xii)}{\geq} 1 - \frac{1}{2} \sqrt{\frac{1}{|\mathcal{A}_n|} e^{\mu_n^2 \tau^2 / 2}} \stackrel{(xiii)}{\geq} \frac{4 - \delta}{4},$$

onde em (xii) usamos a Proposição 8 e a inequação (3.13), e em (xiii) usamos a condição $\mu_n \leq \sqrt{(1/k_n) \log(\delta^2 M_n / 4)}$ e o fato que $\tau^2 / 2 \leq k_n$. Assim, sob essa condição para μ_n , temos que para todo \mathcal{A}_n fixo,

$$\rho_{\mathcal{A}_n}(\mu_n) \geq \frac{1}{2} \mathbb{E}_0 [F(\zeta)] \geq \frac{4 - \delta_n}{8} e^{-\mu_n^2 (k_n - \tau^2 / 2) / 8}.$$

Veja que, além disso pela definição de $t_{c_n}(M_n)$ temos que

$$\begin{aligned}
\frac{1}{2} &\leq \hat{\mathbb{P}}\{\tau \geq t_{c_n}(M_n)\} \\
&= \hat{\mathbb{P}}\{e^{-\mu_n^2 (k_n - \tau^2 / 2) / 8} \geq e^{-\mu_n^2 (k_n - t_{c_n}(M_n)^2 / 2) / 8}\} \\
&\stackrel{(xiv)}{\leq} \hat{\mathbb{E}} \left[e^{-\mu_n^2 (k_n - \tau^2 / 2) / 8} \right] e^{\mu_n^2 (k_n - t_{c_n}(M_n)^2 / 2) / 8},
\end{aligned}$$

onde em (xiv) usamos a desigualdade de Markov. Logo

$$\hat{\mathbb{E}} \left[e^{-\mu_n^2(k_n - \tau^2/2)/8} \right] \geq \frac{1}{2} e^{-\mu_n^2(k_n - t_{\mathcal{C}_n}(M_n)^2/2)/8}.$$

Após aplicar $\hat{\mathbb{E}}$ em

$$\rho_{\mathcal{A}_n}(\mu_n) \geq \frac{4 - \delta}{8} e^{-\mu_n^2(k_n - \tau^2/2)/8},$$

obtemos que

$$\hat{\mathbb{E}} \left[\rho_{\mathcal{A}_n}(\mu_n) \right] \geq \frac{4 - \delta}{16} e^{-\mu_n^2(k_n - t_{\mathcal{C}_n}(M_n)^2/2)/8},$$

onde $t_{\mathcal{C}_n}(M_n)$ é a mediana de τ . Desse modo, queremos μ_n tal que

$$\frac{4 - \delta}{16} e^{-\mu_n^2(k_n - t_{\mathcal{C}_n}(M_n)^2/2)/8} \geq \frac{\sqrt{2\delta - \delta^2}}{2},$$

ou seja, μ_n tal que

$$e^{-\mu_n^2(k_n - t_{\mathcal{C}_n}(M_n)^2/2)/8} \geq \frac{8}{4 - \delta} \sqrt{2\delta - \delta^2}.$$

Seja δ^* o maior valor de $\delta \in (0, 1)$ tal que $\psi(\delta) = \frac{8}{4 - \delta} \sqrt{2\delta - \delta^2} < 1$. Note que $\delta^* > 0$ pois $\psi(0) = 0$ e ψ é função contínua em zero. Logo,

$$\mu_n^2(k_n - t_{\mathcal{C}_n}(M_n)^2/2)/8 \leq \log \left(\frac{4 - \delta^*}{8\sqrt{2\delta^* - (\delta^*)^2}} \right).$$

Para concluir a prova, basta isolar μ_n na inequação acima. □

3.6 Uma análise parcial do teste ótimo contra o teste baseado no máximo

Em todos os exemplos anteriores, cotas superiores para o risco do teste ótimo $R^* = R(f^*)$ foram obtidas através da análise do teste baseado no máximo ou do teste baseado na soma. Como foi mostrado na maioria dos exemplos estudados na Seção 3.3, o risco de um desses testes auxiliares e o risco de Bayes têm o mesmo comportamento assintótico.

Nesta seção, obtemos uma cota superior para a probabilidade $\mathbb{P}_0\{f^*(X) = 1\}$ baseada no valor esperado do máximo de um processo Gaussiano indexado por um conjunto esparsos de \mathcal{C}_n e que pode ser significativamente menor que

o máximo sob toda a classe que aparece na cota do teste baseada no máximo apresentada na Proposição 5. Infelizmente, não temos um cota análoga para a probabilidade $(1/N_n) \sum_{S \in \mathcal{C}_n} \mathbb{P}_S\{f^*(X) = 0\}$ e por isso não é possível obter diretamente uma cota superior para o risco médio do teste da razão de verossimilhança.

Da definição do teste da razão de verossimilhança para o problema de detecção de médias em (2.8) temos que

$$\mathbb{P}_0\{f^*(X) = 1\} = \mathbb{P}_0\{L(X) > 1\} = \mathbb{P}_0\left\{\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} > e^{\frac{k_n \mu_n^2}{2}}\right\}.$$

Como $X_S \leq \max_{S \in \mathcal{C}_n} X_S$, temos que $e^{\mu_n X_S} \leq e^{\mu_n \max_{S \in \mathcal{C}_n} X_S}$ para todo $S \in \mathcal{C}_n$ e assim segue que

$$\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \leq e^{\mu_n \max_{S \in \mathcal{C}_n} X_S}.$$

Portanto,

$$\left\{\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} > e^{\frac{k_n \mu_n^2}{2}}\right\} \subset \left\{e^{\mu_n \max_{S \in \mathcal{C}_n} X_S} > e^{\frac{k_n \mu_n^2}{2}}\right\},$$

e então

$$\mathbb{P}_0\{L(X) > 1\} \leq \mathbb{P}_0\left\{\max_{S \in \mathcal{C}_n} X_S > k_n \mu_n / 2\right\}.$$

Vimos na demonstração da Proposição 5 que $\max_{S \in \mathcal{C}_n} X_S$ é uma função Lipschitz e tem constante de Lipschitz igual a $\sqrt{k_n}$. Desse modo,

$$\begin{aligned} \mathbb{P}_0\left\{\max_{S \in \mathcal{C}} X_S > k_n \mu_n / 2\right\} &= \mathbb{P}_0\left\{\max_{S \in \mathcal{C}_n} X_S - \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} X_S\right] \right. \\ &\quad \left. > k_n \mu_n / 2 - \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} X_S\right]\right\} \\ &\stackrel{(i)}{\leq} \exp\left(\frac{-\left(\frac{k_n \mu_n}{2} - \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} X_S\right]\right)^2}{2k_n}\right), \end{aligned}$$

onde em (i) usamos o Lema 3. Para que $\mathbb{P}_0\{L(X) > 1\} \leq \delta_n$, precisamos que

$$\exp\left(\frac{-\left(\frac{k_n \mu_n}{2} - \mathbb{E}_0\left[\max_{S \in \mathcal{C}_n} X_S\right]\right)^2}{2k_n}\right) \leq \delta_n.$$

Disso segue que μ_n deve ser tal que,

$$\mu_n \geq \frac{2}{k_n} \mathbb{E}_0 \left[\max_{S \in \mathcal{C}_n} X_S \right] + \sqrt{\frac{8}{k_n} \log(1/\delta_n)}.$$

Para entender a diferença entre a performance do teste ótimo f^* e do teste baseado no máximo f_{TM} , precisamos comparar as variáveis aleatórias $\max_{S \in \mathcal{C}_n} X_S$ e $(1/\mu_n) \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right)$.

Dado $t > 0$, dizemos que $\mathcal{G} \subset \mathcal{C}_n$ é uma t -cobertura de \mathcal{C}_n se para qualquer $S \in \mathcal{C}_n$ existe $T \in \mathcal{G}$ tal que $d(T, S) \leq t$ onde $d(T, S)$ está definido em (3.28). Denotamos $\Pi_{\mathcal{C}_n}(t)$ o conjunto de todas as t -coberturas de \mathcal{C}_n e $N(t) = \min_{\mathcal{G} \in \Pi_{\mathcal{C}_n}(t)} |\mathcal{G}|$ é o tamanho da menor t -cobertura de \mathcal{C}_n . Dizemos que $\mathcal{G} \subset \mathcal{C}_n$ é uma t -cobertura minimal de \mathcal{C}_n com cardinalidade $|\mathcal{G}| = N(t)$.

Proposição 12 (Addario-Berry et al. (2010)). *Sejam $\delta_n \in (0, 1)$ e f^* o teste da razão de verossimilhança (teste ótimo). Para que*

$$\mathbb{P}_0\{f^*(X) = 1\} \leq \delta_n,$$

é necessário que

$$\mu_n \geq \frac{2}{k_n} \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] + \sqrt{\frac{8}{k_n} \log(1/\delta_n)},$$

onde \mathcal{A}_n é qualquer $\sqrt{k_n}/2$ -cobertura de \mathcal{C}_n .

Se \mathcal{A}_n é $\sqrt{k_n}/2$ -cobertura minimal de \mathcal{C}_n , então

$$(1/k_n) \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] \leq \sqrt{\frac{2 \log(N_n(\sqrt{k_n}/2))}{k_n}}.$$

Demonstração. Seja $M_{C_n}(\mu_n) = \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right) \right]$. Note que,

$$\begin{aligned}
\mathbb{P}_0\{f^*(X) = 1\} &= \mathbb{P}_0 \left\{ \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} > e^{k_n \mu_n^2 / 2} \right\} \\
&= \mathbb{P}_0 \left\{ \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right) > k_n \mu_n^2 / 2 \right\} \\
&= \mathbb{P}_0 \left\{ \frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right) > k_n \mu_n / 2 \right\} \\
&= \mathbb{P}_0 \left\{ \frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right) - M_{C_n}(\mu_n) \right. \\
&\quad \left. > \frac{k_n \mu_n}{2} - M_{C_n}(\mu_n) \right\} \\
&\stackrel{(ii)}{\leq} \exp \left(\frac{- \left(\frac{k_n \mu_n}{2} - M_{C_n}(\mu_n) \right)^2}{2k_n} \right),
\end{aligned}$$

onde em (ii) usamos o Lema 3. Logo, $\mathbb{P}_0\{f^*(X) = 1\} \leq \delta_n$ desde que

$$\mu_n \geq \frac{2M_{C_n}(\mu_n)}{k_n} + \sqrt{\frac{8}{k_n} \log(1/\delta_n)}.$$

Resta limitar $M_{C_n}(\mu_n)$ e de fato, mostrar que a função $h : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por

$$h(\mathbf{x}) = \frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n x_S} \right),$$

onde $\mathbf{x} = (x_1, \dots, x_n)$ é função Lipschitz e com constante de Lipschitz igual a $\sqrt{k_n}$ para justificar o uso do Lema 3. Veja que,

$$\frac{\partial h(\mathbf{x})}{\partial x_j} = \frac{\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{1}_{\{j \in S\}} e^{\mu_n x_S}}{\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n x_S}} \in (0, 1),$$

logo

$$\|\nabla h(\mathbf{x})\|_2^2 = \sum_{j=1}^n \left(\frac{\partial h(\mathbf{x})}{\partial x_j} \right)^2 \leq \sum_{i=1}^n \frac{\partial h(\mathbf{x})}{\partial x_j} = \frac{\sum_{S \in \mathcal{C}_n} \sum_{j=1}^n \mathbb{1}_{\{j \in S\}} e^{\mu_n x_S}}{\sum_{S \in \mathcal{C}_n} e^{\mu_n x_S}} = k_n,$$

pois $\sum_{j=1}^n \mathbb{1}_{\{j \in S\}} = k_n$. Segue que h é função de Lipschitz com constante de Lipschitz igual a $\sqrt{k_n}$.

Agora, vamos limitar $M_{\mathcal{C}_n}(\mu_n)$. Seja $t \leq \sqrt{2k_n}$ inteiro positivo e denote \mathcal{A}_n a t -cobertura minimal do conjunto \mathcal{C}_n . Por definição, se $\pi(S)$ denota um elemento em \mathcal{A}_n cuja distância a $S \in \mathcal{C}_n$ é minimal, então $d(S, \pi(S)) \leq t$ para todo $S \in \mathcal{C}_n$. Desse modo, veja que

$$\begin{aligned}
M_{\mathcal{C}_n}(\mu_n) &= \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S} \right) \right] \\
&= \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S - \mu_n X_{\pi(S)} + \mu_n X_{\pi(S)}} \right) \right] \\
&= \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S - \mu_n X_{\pi(S)}} e^{\mu_n X_{\pi(S)}} \right) \right] \\
&\leq \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n X_S - \mu_n X_{\pi(S)}} e^{\max_{S \in \mathcal{C}_n} \mu_n X_{\pi(S)}} \right) \right] \\
&= \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n (X_S - X_{\pi(S)})} e^{\max_{S \in \mathcal{A}_n} \mu_n X_S} \right) \right] \\
&= \mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n (X_S - X_{\pi(S)})} \right) \right] + \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right].
\end{aligned}$$

Para limitar o primeiro termo do lado direito da equação acima veja que pela desigualdade de Jensen,

$$\mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n (X_S - X_{\pi(S)})} \right) \right] \leq \frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \mathbb{E}_0 \left[e^{\mu_n (X_S - X_{\pi(S)})} \right] \right).$$

Como, sob \mathbb{P}_0 , vale que

$$X_S - X_T = \sum_{i \in S \setminus T} X_i - \sum_{i \in T \setminus S} X_i \sim \mathcal{N}(0, |S \setminus T| + |T \setminus S|),$$

e

$$|S \setminus T| + |T \setminus S| = d^2(S, T),$$

segue que

$$X_S - X_{\pi(S)} \sim \mathcal{N}(0, d^2(S, \pi(S))).$$

Desse modo, usando a função geradora de momentos da variável aleatória normal e que $d(S, \pi(S)) \leq t$ temos que

$$\mathbb{E}_0 \left[\frac{1}{\mu_n} \log \left(\frac{1}{N_n} \sum_{S \in \mathcal{C}_n} e^{\mu_n(X_S - X_{\pi(S)})} \right) \right] \leq \frac{\mu_n t^2}{2}.$$

Para o segundo termo, temos que pela Proposição 6

$$\mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] \leq \sqrt{2k_n \log(N(t))},$$

onde $N(t) = |\mathcal{A}_n|$ pois \mathcal{A}_n é uma cobertura minimal. Escolhendo $t^2 = k_n/4$, obtemos que

$$\begin{aligned} \mu_n &\geq \frac{2}{k_n} \left(\frac{\mu_n k_n}{8} + \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] \right) + \sqrt{\frac{8}{k_n} \log(1/\delta_n)} \\ &= \frac{\mu_n}{4} + \frac{2}{k_n} \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] + \sqrt{\frac{8}{k_n} \log(1/\delta_n)}. \end{aligned}$$

Logo

$$\frac{3}{4} \mu_n \geq \frac{2}{k_n} \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] + \sqrt{\frac{8}{k_n} \log(1/\delta_n)},$$

e assim

$$\mu_n \geq \frac{2}{k_n} \mathbb{E}_0 \left[\max_{S \in \mathcal{A}_n} X_S \right] + \sqrt{\frac{8}{k_n} \log(1/\delta_n)},$$

concluindo a prova. □

Note que na Proposição 12, podemos substituir o valor esperado $\mathbb{E}_0 [\max_{S \in \mathcal{C}_n} X_S]$ por $\mathbb{E}_0 [\max_{S \in \mathcal{A}_n} X_S]$ obtido pelo teste baseado no máximo onde agora o máximo é tomado sob um subconjunto potencialmente muito menor $\mathcal{A}_n \subset \mathcal{C}_n$. Nos casos em que $\mathbb{E}_0 [\max_{S \in \mathcal{A}_n} X_S] \ll \mathbb{E}_0 [\max_{S \in \mathcal{C}_n} X_S]$, é intuitivo pensar que o teste baseado no máximo não é ótimo. Entretanto, para verificar essa conjectura seria necessário fazer uma análise similar para a probabilidade $(1/N_n) \sum_{S \in \mathcal{C}_n} \mathbb{P}_S \{f^*(X) = 0\}$.

Capítulo 4

Detecção de correlações

Existe uma relação próxima entre o problema de detecção de médias e o problema de detecção de correlações que é estabelecida pelo Teorema de representação de (Berman (1962)), aqui apresentado no caso de variáveis aleatórias Gaussianas.

Lema 8. *Sejam X_1, \dots, X_{k_n} variáveis aleatórias com distribuição normal padrão tais que $\text{Cov}(X_i, X_j) = \rho_n$ para $i \neq j$ e $1 \leq i \leq k_n$. Então existem variáveis aleatórias com distribuição normal padrão independentes e identicamente distribuídas denotadas por U, U_1, \dots, U_{k_n} tais que $X_i = \sqrt{\rho_n}U + \sqrt{1 - \rho_n}U_i$ para todo i .*

Demonstração. Seja para todo $i \in \{1, \dots, k_n\}$, $X_i = \sqrt{\rho_n}U + \sqrt{1 - \rho_n}U_i$ com U, U_1, \dots, U_{k_n} variáveis aleatórias normais padrão independentes e identicamente distribuídas. Logo, para todo $t \in \mathbb{R}$ temos que

$$\begin{aligned} M_{X_i}(t) &= \mathbb{E} \left[e^{t\sqrt{\rho_n}U + t\sqrt{1-\rho_n}U_i} \right] \\ &\stackrel{(i)}{=} e^{\frac{t^2\rho_n}{2} + \frac{t^2(1-\rho_n)}{2}} \\ &= e^{\frac{t^2}{2}}, \end{aligned}$$

onde em (i) usamos que U, U_1, \dots, U_{k_n} são variáveis aleatórias independentes com distribuição normal padrão. Desse modo, $X_i \sim \mathcal{N}(0, 1)$ para todo $i \in \{1, \dots, k_n\}$.

Além disso, para $i \neq j$ temos que

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \mathbb{E} \left[\left(\sqrt{\rho_n}U + \sqrt{1 - \rho_n}U_i \right) \left(\sqrt{\rho_n}U + \sqrt{1 - \rho_n}U_j \right) \right] \\ &= \rho_n \mathbb{E} [U^2] = \rho_n. \end{aligned}$$

□

Desse modo, condicionando em $\{U = u\}$, o problema de detecção de correlações se reduz ao problema de detectar subconjuntos de variáveis Gaussianas com média não nula (igual a $\sqrt{\rho_n}u$) e com variância igual a $1 - \rho_n$ (ao invés de 1). Essa observação será muito útil na subseção 4.1.2. Quando U é aleatório, o problema é semelhante ao de detectar a presença de um ruído Gaussiano aditivo em um processo Gaussiano (aqui igual a $\sqrt{\rho_n}U$ para $i \in S$ e igual a 0 caso contrário). No entanto, a configuração típica dos processos Gaussianos assume que os efeitos afetam todas as partes do sinal (Kailath and Poor (1998)). No nosso contexto, o sinal (o subconjunto de variáveis correlacionadas) será esparsos. Em certo sentido, o ρ_n no problema de detecção de correlações desempenha papel análogo ao μ_n no problema de detecção de médias.

4.1 Cotas inferiores

Nesta seção investigamos cotas inferiores para o risco de Bayes. Primeiramente, consideramos o caso em que \mathcal{C}_n contém apenas um conjunto de variáveis anômalas. Depois, obtemos uma cota inferior geral para o risco de Bayes baseada no Lema 8 que tem importantes implicações e que será usada nos nossos exemplos. A saber, essa cota é a análoga a da Proposição 8 para o problema de detecção de médias.

4.1.1 Caso $N_n = 1$

Considere o caso em que \mathcal{C}_n contém apenas um conjunto, denotado por $S = \{1, \dots, k_n\}$. Nesse caso, a hipótese alternativa é simples e o teste da razão de verossimilhança definido em (2.11) pode ser expressado por

$$f^*(X) = 0 \text{ se e somente se } X^T(\mathbf{I}_n - A_S^{-1})X \leq \log(\det(A_S)),$$

onde a matriz A_S é definida em (2.3). O próximo resultado fornece a distribuição da variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ sob \mathbb{P}_0 e sob \mathbb{P}_S , onde \mathbb{P}_0 é a medida de probabilidade sob a qual $X \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$ e \mathbb{P}_S é a medida de probabilidade sob a qual $X \sim \mathcal{N}(\mathbf{0}_n, A_S)$. Esse resultado é importante e será usado ao longo desse capítulo.

Lema 9 (Arias-Castro et al. (2012)). *Sob \mathbb{P}_0 , a variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ tem distribuição*

$$-\frac{\rho_n}{1 - \rho_n} \chi_{k_n-1}^2 + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)} \chi_1^2,$$

e sob \mathbb{P}_S , tem distribuição de

$$-\rho_n \chi_{k_n-1}^2 + \rho_n(k_n - 1)\chi_1^2,$$

onde χ_1^2 e $\chi_{k_n-1}^2$ denotam variáveis aleatórias independentes com distribuição qui-quadrado com 1 e $k_n - 1$ graus de liberdade, respectivamente.

Demonstração. Veja que se B é matriz tal que $BB^T = \Sigma$ e $Y \sim \mathcal{N}(0, \mathbf{I}_n)$, logo $BY \sim \mathcal{N}(0, \Sigma)$ (ver definição 7.1 em (James (1996))). Como A_S é matriz simétrica, pelo teorema espectral para matrizes reais simétricas temos que existe matriz diagonal D e matriz ortogonal U tais que $A_S = UDU^T$ e portanto $A_S^{1/2} = UD^{1/2}U^T$.

Desse modo, segue que $A_S^{1/2} (A_S^{1/2})^T = A_S$ e $A_S^{1/2}Y \sim \mathcal{N}(0, A_S)$. Dessas observações, temos que sob \mathbb{P}_0 , a variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ segue a mesma distribuição que $Y^T(\mathbf{I}_n - A_S^{-1})Y$ e sob \mathbb{P}_S , a variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ segue a mesma distribuição que $(A_S Y)^T(\mathbf{I}_n - A_S^{-1})A_S Y$.

Portanto sob \mathbb{P}_0 , a variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ segue a mesma distribuição que $Y^T(\mathbf{I}_n - A_S^{-1})Y$ e sob \mathbb{P}_S , a variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ segue a mesma distribuição que $Y^T(A_S - \mathbf{I}_n)Y$.

Veja que como B é matriz simétrica, temos pelo teorema espectral para matrizes simétricas temos que a matriz B pode ser escrita como $B = Q^T \Lambda Q$, onde Q é matriz ortogonal de autovetores e Λ é matriz diagonal de autovalores. Desse modo, vemos que

$$\begin{aligned} Y^T B Y &= Y^T (Q^T \Lambda Q) Y \\ &= (QY)^T \Lambda (QY), \end{aligned}$$

que tem a mesma distribuição que a variável aleatória $Y^T \Lambda Y$ pois como Q é ortogonal temos que $QY \sim \mathcal{N}(0, \mathbf{I}_n)$. Assim, a variável aleatória $Y^T B Y$ tem a mesma distribuição que $\sum_{i=1}^n \lambda_i Y_i^2$.

Vamos calcular os autovalores de A_S . Note que podemos reescrever a matriz A_S como

$$\begin{aligned} A_S &= \left[\begin{array}{c|c} M(\rho_n) & \mathbf{0}_{k_n \times n-k_n} \\ \hline \mathbf{0}_{n-k_n \times k_n} & \mathbf{I}_{n-k_n} \end{array} \right]_{n \times n} \\ &= \mathbf{I}_n + \rho_n \left[\begin{array}{c|c} \mathbf{1}_{k_n} & \mathbf{0}_{k_n \times n-k_n} \\ \hline \mathbf{0}_{n-k_n \times k_n} & \mathbf{0}_{n-k_n \times n-k_n} \end{array} \right] - \rho_n \left[\begin{array}{c|c} \mathbf{I}_{k_n} & \mathbf{0}_{k_n \times n-k_n} \\ \hline \mathbf{0}_{n-k_n \times k_n} & \mathbf{0}_{n-k_n \times n-k_n} \end{array} \right], \end{aligned}$$

onde

$$M(\rho_n) = \begin{bmatrix} 1 & & \rho_n \\ & \ddots & \\ \rho_n & & 1 \end{bmatrix},$$

onde $\mathbf{1}_{k_n}$ é matriz de dimensão $k_n \times k_n$ cujas entradas são todas iguais a 1, $\mathbf{0}_{a \times b}$ é matriz de dimensão $a \times b$ cujas entradas são todas iguais à zero e $\mathbf{0}_a$ é a matriz de dimensão $a \times a$ cujas entradas são todas iguais a zero.

Quero resolver $(A_S - \lambda \mathbf{I}_n)v = 0$, que é equivalente à

$$\left((1 - \lambda)\mathbf{I}_n + \rho_n \left[\begin{array}{c|c} \mathbf{1}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] - \rho_n \left[\begin{array}{c|c} \mathbf{I}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] \right) v = 0.$$

Tomando $\lambda = 1$ temos que a igualdade acima é equivalente à $\rho_n(v_1 + \dots + v_{k_n}) = \rho_n v_1 = \dots = \rho_n v_{k_n}$, o que implica $v_1 = \dots = v_{k_n} = 0$. Assim, $\lambda = 1$ é autovalor com multiplicidade $n - k_n$.

Agora, tomando $\lambda = 1 - \rho_n$ temos que

$$\left(\rho_n \mathbf{I}_n + \rho_n \left[\begin{array}{c|c} \mathbf{1}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] - \rho_n \left[\begin{array}{c|c} \mathbf{I}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] \right) v = 0,$$

que é equivalente à

$$\left(\rho_n \left[\begin{array}{c|c} \mathbf{1}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] + \rho_n \left[\begin{array}{c|c} \mathbf{0}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{I}_{n - k_n} \end{array} \right] \right) v = 0.$$

Logo, $v_{k_n+1} = \dots = v_n = 0$ e $v_1 + \dots + v_{k_n} = 0$. Assim, $\lambda = 1 - \rho_n$ é autovalor com multiplicidade $k_n - 1$.

Por fim, tomando $\lambda = 1 + \rho_n(k_n - 1)$ temos que

$$\left(-\rho_n(k_n - 1)\mathbf{I}_n + \rho_n \left[\begin{array}{c|c} \mathbf{1}_{k_n \times k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] - \rho_n \left[\begin{array}{c|c} \mathbf{I}_{k_n} & \mathbf{0}_{k_n \times n - k_n} \\ \hline \mathbf{0}_{n - k_n \times k_n} & \mathbf{0}_{n - k_n} \end{array} \right] \right) v = 0,$$

que é equivalente à $v_{k_n+1} = \dots = v_n = 0$ e $\rho_n(v_1 + \dots + v_{k_n}) = \rho_n k_n v_1 = \dots = \rho_n k_n v_{k_n}$, o que implica em $v_1 = \dots = v_{k_n}$. Assim, $\lambda = 1 + \rho_n(k_n - 1)$ é autovalor com multiplicidade 1.

Vamos determinar a distribuição sob \mathbb{P}_S . Veja que os autovalores de

$$(A_S - \mathbf{I}_n)v = A_S v - \mathbf{I}_n v = \lambda v - v = (\lambda - 1)v.$$

Logo $(A_S - \mathbf{I}_n)$ tem autovalores 0 com multiplicidade $n - k_n$, $-\rho$ com multiplicidade $k_n - 1$ e $\rho_n(k_n - 1)$ com multiplicidade 1. Segue que $Y^T(A_S - \mathbf{I}_n)Y \sim -\rho_n \chi_{k_n-1}^2 + \rho_n(k_n - 1)\chi_1^2$ onde χ_1^2 e $\chi_{k_n-1}^2$ denotam variáveis aleatórias independentes qui-quadrado com 1 e $k_n - 1$ graus de liberdade, respectivamente.

Vamos determinar a distribuição sob \mathbb{P}_0 . Como $A_S v = \lambda v$, logo $v = \lambda A_S^{-1} v$ e então $A_S^{-1} v = \frac{v}{\lambda}$. Assim,

$$(\mathbf{I}_n - A_S^{-1})v = v - \frac{v}{\lambda} = \frac{\lambda - 1}{\lambda}v,$$

onde λ é autovalor de A_S . Logo, $(\mathbf{I}_n - A_S^{-1})$ tem autovalores 0 com multiplicidade $n - k_n$, $\frac{-\rho_n}{1 - \rho_n}$ com multiplicidade $k_n - 1$ e $\frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}$ com multiplicidade 1. Segue que $Y^T(\mathbf{I}_n - A_S^{-1})Y$ tem distribuição $-\frac{\rho_n}{1 - \rho_n}\chi_{k_n - 1}^2 + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}\chi_1^2$. \square

Na seguinte proposição analisamos o risco de Bayes.

Proposição 13 (Arias-Castro et al. (2012)). *Se \mathcal{C}_n tem um único conjunto, então $\lim_{k_n \rightarrow \infty} R^* = 0$ se e somente se $\rho_n k_n \rightarrow \infty$. Analogamente, $\lim_{k_n \rightarrow \infty} R^* = 1$ se e somente se $\rho_n k_n \rightarrow 0$.*

Demonstração. Suponha que $\rho_n k_n \rightarrow \infty$. É suficiente encontrarmos τ_{k_n} tal que $\mathbb{P}_0\{X^T(\mathbf{I}_n - A_S^{-1})X \geq \tau_{k_n}\} \rightarrow 0$ e $\mathbb{P}_S\{X^T(\mathbf{I}_n - A_S^{-1})X < \tau_{k_n}\} \rightarrow 0$ pois R^* por definição é o que tem menor risco dentre todos os testes. Como queremos provar o resultado para quando $k_n \rightarrow \infty$, podemos substituir, por conveniência, $k_n - 1$ por k_n . Desse modo, pelo Lema 9 temos que sob \mathbb{P}_0 ,

$$X^T(\mathbf{I}_n - A_S^{-1})X \sim -\frac{\rho_n}{1 - \rho_n}\chi_{k_n}^2 + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2.$$

Como $\chi_{k_n}^2 \geq 0$ e $\frac{1}{1 - \rho_n} > 1$, logo $\frac{-\rho_n \chi_{k_n}^2}{1 - \rho_n} < -\rho_n \chi_{k_n}^2$. Desse modo,

$$\begin{aligned} \mathbb{P}\left\{-\frac{\rho_n}{1 - \rho_n}\chi_{k_n}^2 + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2 \geq \tau_{k_n}\right\} &\leq \mathbb{P}\left\{-\rho_n \chi_{k_n}^2 + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2 \geq \tau_{k_n}\right\} \\ &= \mathbb{P}\left\{-\rho_n(\chi_{k_n}^2 - k_n) - \rho_n k_n + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2 \geq \tau_{k_n}\right\} \\ &= \mathbb{P}\left\{-\rho_n(\chi_{k_n}^2 - k_n) + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2 \geq \tau_{k_n} + \rho_n k_n\right\} \\ &\leq \mathbb{P}\left\{\rho_n|\chi_{k_n}^2 - k_n| + \frac{\rho_n k_n}{1 + \rho_n k_n}\chi_1^2 \geq \tau_{k_n} + \rho_n k_n\right\}. \end{aligned}$$

Seja t_{k_n} sequência tal que $t_{k_n} \rightarrow \infty$ e $\tau_{k_n} + \rho_n k_n = \rho_n \sqrt{k_n} t_{k_n} + t_{k_n}$. Note que,

$$\left\{ \rho_n |\chi_{k_n}^2 - k_n| + \frac{\rho_n k_n}{1 + \rho_n k_n} \chi_1^2 \geq \rho_n \sqrt{k_n} t_{k_n} + t_{k_n} \right\} \subseteq \left\{ \rho_n |\chi_{k_n}^2 - k_n| \geq \rho_n \sqrt{k_n} t_{k_n} \right\} \cup \left\{ \frac{\rho_n k_n}{1 + \rho_n k_n} \chi_1^2 \geq t_{k_n} \right\}. \quad (4.1)$$

Pela desigualdade de Chebyshev,

$$\mathbb{P} \left\{ \rho_n |\chi_{k_n}^2 - k_n| \geq \rho_n \sqrt{k_n} t_{k_n} \right\} \leq \frac{2}{t_{k_n}^2} \rightarrow 0, \quad (4.2)$$

quando $n \rightarrow \infty$ e como $\frac{\rho_n k_n}{1 + \rho_n k_n} \leq 1$, logo

$$\mathbb{P} \left\{ \frac{\rho_n k_n}{1 + \rho_n k_n} \chi_1^2 \geq t_{k_n} \right\} \leq \mathbb{P} \left\{ \chi_1^2 \geq t_{k_n} \right\} \rightarrow 0, \quad (4.3)$$

quando $n \rightarrow \infty$. De (4.1), (4.2), (4.3) e da cota da união temos que

$$\mathbb{P}_0 \left\{ X^T (\mathbf{I}_n - A_S^{-1}) X \geq \tau_{k_n} \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$. Pelo Lema 9, sob \mathbb{P}_S , temos que

$$X^T (\mathbf{I}_n - A_S^{-1}) X \sim -\rho_n \chi_{k_n}^2 + \rho_n k_n \chi_1^2.$$

De maneira análoga vemos que,

$$\mathbb{P} \left\{ -\rho_n \chi_{k_n}^2 + \rho_n k_n \chi_1^2 < \eta_{k_n} \right\} = \mathbb{P} \left\{ -\rho_n (\chi_{k_n}^2 - k_n) + \rho_n k_n \chi_1^2 < \eta_{k_n} + \rho_n k_n \right\}.$$

Seja $\eta_{k_n} + \rho_n k_n = -\rho_n \sqrt{k_n} t_{k_n} + \rho_n k_n t_{k_n}^{-1}$. Logo,

$$\begin{aligned} \mathbb{P} \left\{ -\rho_n \chi_{k_n}^2 + \rho_n k_n \chi_1^2 < \eta_{k_n} \right\} &\leq \mathbb{P} \left\{ -\rho_n (\chi_{k_n}^2 - k_n) < -\rho_n \sqrt{k_n} t_{k_n} \right\} \\ &\quad + \mathbb{P} \left\{ \rho_n k_n \chi_1^2 < \rho_n k_n t_{k_n}^{-1} \right\} \\ &\leq \mathbb{P} \left\{ |\chi_{k_n}^2 - k_n| \geq \sqrt{k_n} t_{k_n} \right\} + \mathbb{P} \left\{ \chi_1^2 < t_{k_n}^{-1} \right\}, \end{aligned}$$

que tende à zero quando $t_{k_n} \rightarrow \infty$. Desse modo, segue que

$$\mathbb{P}_S \left\{ X^T (\mathbf{I}_n - A_S^{-1}) X < \eta_{k_n} \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$. Veja que quando k_n é grande o suficiente, $\tau_{k_n} < \eta_{k_n}$ e isso implica que

$$\mathbb{P}_S \left\{ X^T (\mathbf{I}_n - A_S^{-1}) X < \tau_{k_n} \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$.

Agora, suponha que $\rho_n k_n$ é limitado. Assim, queremos mostrar que $\lim_{k_n \rightarrow \infty} R^* \neq 0$ quando $n \rightarrow \infty$. Logo, $\rho_n k_n \rightarrow L < \infty$ quando $n \rightarrow \infty$ e $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$.

Denote $Z_n = -\frac{\rho_n}{1-\rho_n} \chi_{k_n}^2 + \frac{\rho_n k_n}{1+\rho_n k_n} \chi_1^2$. Note que pela Lei dos Grandes Números,

$$-\frac{\rho_n k_n}{(1-\rho_n)k_n} \chi_{k_n}^2 \xrightarrow[n \rightarrow \infty]{q.c.} -L.$$

Além disso,

$$\frac{\rho_n k_n}{1+\rho_n k_n} \chi_1^2 \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \frac{L}{1+L} \chi_1^2.$$

Portanto, pelo Teorema de Slutsky

$$Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} -L + \frac{L}{1+L} \chi_1^2.$$

Denote $W_n = -\rho_n \chi_{k_n}^2 + \rho_n k_n \chi_1^2$. De modo análogo temos que

$$W_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} -L + L \chi_1^2.$$

Tome $t_n = k_n \log(1 - \rho_n) + \log(1 + \rho_n k_n)$. Veja que

$$t_n = \rho_n k_n \frac{\log(1 - \rho_n)}{\rho_n} + \log(1 + \rho_n k_n).$$

Como $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$, logo $t_n \rightarrow -L + \log(1 + L)$ quando $n \rightarrow \infty$. Desse modo,

$$\begin{aligned} R^* &= \mathbb{P}\{Z_n \geq t_n\} + \mathbb{P}\{W_n < t_n\} \\ &\rightarrow \mathbb{P}\left\{-L + \frac{L}{1+L} \chi_1^2 \geq -L + \log(1 + L)\right\} \\ &\quad + \mathbb{P}\left\{-L + L \chi_1^2 \geq -L + \log(1 + L)\right\} \\ &= \mathbb{P}\left\{\chi_1^2 \geq \frac{1+L}{L} \log(1 + L)\right\} \\ &\quad + \mathbb{P}\left\{\chi_1^2 < \frac{\log(1 + L)}{L}\right\} \\ &> 0, \end{aligned}$$

quando $n \rightarrow \infty$, onde usamos o Teorema de Slutsky.

A prova do outro caso é similar. □

4.1.2 Cota inferior baseada em pares sobrepostos

Nesta subsecção, independentemente da classe \mathcal{C}_n escolhida, obtemos cotas inferiores para o risco de Bayes. Assim como no problema de detecção de médias, a medida de complexidade importante é dada em termos da função geradora de momentos do tamanho da sobreposição de dois elementos escolhidos ao acaso e de maneira independente em \mathcal{C}_n . Usamos o Lema 8 para fazer uma conexão entre o problema de detecção de médias e o problema de detecção de correlações.

Teorema 6 (Arias-Castro et al. (2012)). *Para qualquer classe \mathcal{C}_n e qualquer $a > 0$,*

$$R^* \geq \mathbb{P}\{|\mathcal{N}(0, 1)| \leq a\} \left(1 - \frac{1}{2} \sqrt{\mathbb{E} [\exp(\nu_a Z)]} - 1\right),$$

onde $\nu_a := \frac{\rho_n}{1+\rho_n} a^2 - \frac{1}{2} \log(1 - \rho_n^2)$ e $Z = |S \cap S'|$ com S e S' escolhidos ao acaso e de maneira independente em \mathcal{C}_n . Em particular, tomando $a = 1$,

$$R^* \geq 0.6 - 0.3 \sqrt{\mathbb{E} [\exp(\nu_1 Z)]} - 1,$$

onde $\nu_1 = \frac{\rho_n}{1+\rho_n} - \frac{1}{2} \log(1 - \rho_n^2)$.

Demonstração. Pelo Lema 8, o vetor X pode ser representado como

$$X_i = \begin{cases} U_i, & \text{se } i \notin S, \\ \sqrt{\rho_n} U + \sqrt{1 - \rho_n} U_i, & \text{se } i \in S, \end{cases}$$

onde U, U_1, \dots, U_n são variáveis aleatórias normais padrão independentes e identicamente distribuídas.

Considere agora a hipótese alternativa $H_1(u)$ definida como H_1 dado $U = u$. Sejam $R(f), L, f^*$ (respectivamente $R_u(f), L_u, f_u^*$) o risco médio do teste f , a razão de verossimilhança e o teste ótimo para H_0 contra H_1 (respectivamente H_0 contra $H_1(u)$). Lembre que para cada $u \in \mathbb{R}$, $R_u(f_u^*) \leq R_u(f^*)$ pela otimalidade de f_u^* para H_0 contra $H_1(u)$. Então, condicionando em U ,

$$\begin{aligned} R^* &= R(f^*) \\ &= 1 - \frac{1}{2} \mathbb{E}_0 [|L(X) - 1|] \\ &= \mathbb{E} \left[1 - \frac{1}{2} \mathbb{E}_0 [|L(X) - 1| | U] \right] \\ &= \mathbb{E}_U [R_U(f^*)] \\ &\geq \mathbb{E}_U [R_U(f_u^*)] \\ &= 1 - \frac{1}{2} \mathbb{E}_U [\mathbb{E}_0 [|L_U(X) - 1|]], \end{aligned}$$

onde \mathbb{E}_U é a esperança com respeito a $U \sim \mathcal{N}(0, 1)$. Como $\mathbb{E}_0 [|L_u(X) - 1|] \leq 2$ para todo u , logo

$$\begin{aligned} \mathbb{E}_U \left[\mathbb{E}_0 [|L_U(X) - 1|] \right] &= \mathbb{E}_U \left[(\mathbb{1}_{\{|U| > a\}} + \mathbb{1}_{\{|U| \leq a\}}) \mathbb{E}_0 [|L_U(X) - 1|] \right] \\ &\leq 2\mathbb{E}_U [\mathbb{1}_{\{|U| > a\}}] + \mathbb{E}_U \left[\mathbb{1}_{\{|U| \leq a\}} \max_{u \in [-a, a]} \mathbb{E}_0 [|L_u(X) - 1|] \right] \\ &= 2\mathbb{P}\{|U| > a\} + \mathbb{P}\{|U| \leq a\} \max_{u \in [-a, a]} \mathbb{E}_0 [|L_u(X) - 1|]. \end{aligned}$$

Desse modo,

$$\begin{aligned} 1 - \frac{1}{2}\mathbb{E}_U \left[\mathbb{E}_0 [|L_U(X) - 1|] \right] &\geq \mathbb{P}\{|U| \leq a\} \left(1 - \frac{1}{2} \max_{u \in [-a, a]} \mathbb{E}_0 [|L_u(X) - 1|] \right) \\ &\stackrel{(i)}{\geq} \mathbb{P}\{|U| \leq a\} \left(1 - \frac{1}{2} \max_{u \in [-a, a]} \sqrt{\mathbb{E}_0 [(L_u(X) - 1)^2]} \right) \\ &\stackrel{(ii)}{=} \mathbb{P}\{|U| \leq a\} \left(1 - \frac{1}{2} \max_{u \in [-a, a]} \sqrt{\mathbb{E}_0 [L_u^2(X) - 1]} \right), \end{aligned}$$

onde em (i) usamos a desigualdade de Cauchy-Schwarz e em (ii) usamos que $\mathbb{E}_0 [L_u(X)] = 1$. Como sob $H_1(u)$, X_1, \dots, X_n são variáveis aleatórias independentes tais que

$$X_i \sim \begin{cases} \mathcal{N}(0, 1), & \text{se } i \notin S, \\ \mathcal{N}(\sqrt{\rho_n}u, (1 - \rho_n)), & \text{se } i \in S, \end{cases}$$

logo

$$\begin{aligned} L_u(x) &= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \frac{\exp(\sum_{i=1}^n x_i^2/2)}{(1 - \rho_n)^{k_n/2}} \exp \left(-\frac{1}{2} \sum_{i \in S} \frac{(x_i - \sqrt{\rho_n}u)^2}{1 - \rho_n} - \frac{1}{2} \sum_{i \notin S} x_i^2 \right) \\ &= \frac{1}{N_n} \sum_{S \in \mathcal{C}_n} \frac{1}{(1 - \rho_n)^{k_n/2}} \exp \left(\frac{1}{2} \sum_{i \in S} \left(x_i^2 - \frac{(x_i - \sqrt{\rho_n}u)^2}{1 - \rho_n} \right) \right). \end{aligned}$$

Assim,

$$\begin{aligned}
\mathbb{E} [L_u^2(X)] &= \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \frac{1}{(1 - \rho_n)^{k_n}} \mathbb{E}_0 \left[\exp \left(\sum_{S \cap S'} \left(X_i^2 - \frac{(X_i - \sqrt{\rho_n}u)^2}{1 - \rho_n} \right) \right. \right. \\
&\quad \left. \left. + \sum_{i \in S \Delta S'} \left(\frac{X_i^2}{2} - \frac{(X_i - \sqrt{\rho_n}u)^2}{2(1 - \rho_n)} \right) \right) \right] \\
&= \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \frac{1}{(2\pi)^{n/2} (1 - \rho_n)^{k_n}} \\
&\quad \times \int_{\mathbb{R}^n} \exp \left(\sum_{i \in S \cap S'} \left(\frac{x_i^2}{2} - \frac{(x_i - \sqrt{\rho_n}u)^2}{1 - \rho_n} \right) \right. \\
&\quad \left. - \sum_{i \in S \Delta S'} \frac{(x_i - \sqrt{\rho_n}u)^2}{2(1 - \rho_n)} - \sum_{i \notin S \cup S'} \frac{x_i^2}{2} \right) dx,
\end{aligned}$$

onde $S \Delta S' = (S \setminus S') \cup (S' \setminus S)$. Pode-se verificar que

$$\frac{x_i^2}{2} - \frac{(x_i - \sqrt{\rho_n}u)^2}{1 - \rho_n} = \frac{\rho_n u^2}{1 + \rho_n} - \frac{1 + \rho_n}{2(1 - \rho_n)} \left(x_i - \frac{2\sqrt{\rho_n}u}{1 + \rho_n} \right)^2.$$

Logo,

$$\begin{aligned}
\mathbb{E}_0 [L_u^2(X)] &= \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \frac{\exp\left((\rho_n u^2 / (1 + \rho_n)) |S \cap S'|\right)}{(2\pi)^{n/2} (1 - \rho_n)^{k_n}} \\
&\quad \times \int_{\mathbb{R}^n} \exp\left(-\sum_{i \in S \cap S'} \frac{1 + \rho_n}{2(1 - \rho_n)} \left(x_i - \frac{2\sqrt{\rho_n} u}{1 + \rho_n}\right)^2\right. \\
&\quad \left.- \sum_{i \in S \Delta S'} \frac{(x_i - \sqrt{\rho_n} u)^2}{2(1 - \rho_n)} - \sum_{i \notin S \cup S'} \frac{x_i^2}{2}\right) dx \\
&\stackrel{(iii)}{=} \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \frac{\exp\left((\rho_n u^2 / (1 + \rho_n)) |S \cap S'|\right)}{(1 - \rho_n)^{k_n}} \left(\frac{1 - \rho_n}{1 + \rho_n}\right)^{|S \cap S'|/2} \\
&\quad \times (1 - \rho_n)^{k_n - |S \cap S'|} \\
&= \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \exp\left((\rho_n u^2 / (1 + \rho_n)) |S \cap S'|\right) \\
&\quad \times \exp\left(-\frac{|S \cap S'|}{2} \log(1 - \rho_n^2)\right) \\
&= \frac{1}{N_n^2} \sum_{S, S' \in \mathcal{C}_n} \exp\left(\left(\frac{\rho_n u^2}{1 + \rho_n} - \frac{1}{2} \log(1 - \rho_n^2)\right) |S \cap S'|\right) \\
&\stackrel{(iv)}{=} \mathbb{E} [\exp(\nu_u Z)],
\end{aligned}$$

onde em (iii) completamos a densidade da normal e em (iv) denotamos $\nu_u = \frac{\rho_n}{1 + \rho_n} u^2 - \frac{1}{2} \log(1 - \rho_n^2)$ e $Z = |S \cap S'|$ com S e S' escolhidos ao acaso e de maneira independente em \mathcal{C}_n . Como $u \in [-a, a]$, logo

$$\nu_u = \frac{\rho_n}{1 + \rho_n} u^2 - \frac{1}{2} \log(1 - \rho_n^2) \leq \frac{\rho_n}{1 + \rho_n} a^2 - \frac{1}{2} \log(1 - \rho_n^2) = \nu_a.$$

Assim, $\mathbb{E}_0 [L_u^2(X)] \leq \mathbb{E}_0 [L_a^2(X)]$ para todo $u \in [-a, a]$ e portanto

$$R^* \geq \mathbb{P}\{|\mathcal{N}(0, 1)| \leq a\} \left(1 - \frac{1}{2} \sqrt{\mathbb{E} [\exp(\nu_a Z)]} - 1\right).$$

□

4.2 Exemplos

Nesta seção, vamos aplicar o Teorema 6 a alguns exemplos já apresentados nessa dissertação. Esse teorema reduz o problema de obter cotas inferiores

para o risco de Bayes ao problema de obter cotas superiores para a a função geradora de momentos de Z . Para tratar desse último problema, vamos usar os resultados apresentados na Seção 3.3.

4.2.1 Conjuntos disjuntos

Nesta subseção, considere o caso em que \mathcal{C}_n contém N_n conjuntos disjuntos de tamanho k_n .

Corolário 2 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n a classe contendo N_n conjuntos disjuntos de tamanho k_n . Se*

$$\nu_1 \leq \frac{\log N_n}{k_n},$$

então o risco de Bayes satisfaz $R^ \geq 0.3$, e $R^* \rightarrow 1$ se $\rho_n \ll \min(1, \log(N_n)/k_n)$ ou $(1 - \rho_n)(N_n)^{2/k_n} \rightarrow \infty$ quando $n \rightarrow \infty$.*

Observação 4. Note que a cota inferior de R^* também depende de a_n . Além disso, por um lado note que para $R^* \rightarrow 1$ quando $n \rightarrow \infty$ é preciso que $a_n \rightarrow \infty$. Por outro lado, se $a_n \rightarrow \infty$ muito rápido a função geradora de momentos avaliado em ν_{a_n} pode não ir para 1. Como veremos na prova do Corolário 2 se $\rho_n \ll \min(1, \log(N_n)/k_n)$ ou $(1 - \rho_n)(N_n)^{2/k_n} \rightarrow \infty$ quando $n \rightarrow \infty$, então é possível encontrar $a_n \rightarrow \infty$ de modo que a função geradora de momentos avaliada em ν_{a_n} converge para 1 fazendo com que o $R^* \rightarrow 1$.

Demonstração. Do Exemplo 3.3.1 temos que

$$\mathbb{E} \left[e^{\nu_1 Z} \right] - 1 = \frac{1}{N_n} \left(e^{\nu_1 k_n} - 1 \right) \leq \frac{1}{N_n} e^{\nu_1 k_n},$$

que é limitado por 1 quando $\nu_1 \leq \log N_n/k_n$. Segue pelo Teorema 6 que $R^* \geq 0.3$.

Queremos mostrar que $R^* \rightarrow 1$ quando $n \rightarrow \infty$. Pelo Teorema 6, isso ocorre quando $a_n \rightarrow \infty$ e $\frac{1}{N_n} e^{\nu_{a_n} k_n} \rightarrow 1$. Desse modo, precisamos mostrar que

$$a_n \rightarrow \infty \quad \text{e} \quad \nu_{a_n} k_n - \log N_n \rightarrow -\infty.$$

Suponha que $\rho_n \ll \min(1, \log(N_n)/k_n)$. Logo $\rho_n \rightarrow 0$ e $\frac{\rho_n}{\log(N_n/k_n)} \rightarrow 0$ quando

$n \rightarrow \infty$. Veja que

$$\begin{aligned} \frac{\nu_{a_n} k_n}{\rho_n k_n a_n^2} &= \frac{1}{1 + \rho_n} - \frac{\log(1 - \rho_n^2)}{2\rho_n a_n^2} \\ &\stackrel{(i)}{=} \frac{1}{1 + \rho_n} - \frac{1}{2\rho_n a_n^2} \sum_{i=1}^{\infty} \left(\frac{-\rho_n^{i+1}}{i} \right) \\ &= \frac{1}{1 + \rho_n} - \frac{1}{2a_n^2} \sum_{i=1}^{\infty} \left(\frac{-\rho_n^i}{i} \right), \end{aligned}$$

onde em (i) usamos a expansão de Taylor de $\log(1 - \rho_n^2)$. Logo, quando $n \rightarrow \infty$, temos que $\nu_{a_n} k_n / \rho_n k_n a_n^2 \rightarrow 1$ desde que $a_n \rightarrow \infty$. Escolhemos $a_n \rightarrow \infty$ quando $n \rightarrow \infty$ de modo que $\frac{\rho_n k_n a_n^2}{\log N_n} \rightarrow 0$. Note que

$$\begin{aligned} \nu_{a_n} k_n - \log N_n &= -\log N_n \left(1 - \frac{\nu_{a_n} k_n}{\log N_n} \right) \\ &= -\log N_n \left(1 + \frac{k_n \log(1 - \rho_n^2)}{2 \log N_n} - \frac{\rho_n k_n a_n^2}{(1 + \rho_n) \log N_n} \right) \\ &= -\log N_n \left(1 - \frac{\rho_n k_n a_n^2}{\log N_n} \left(\frac{1}{1 + \rho_n} - \frac{\log(1 - \rho_n^2)}{2\rho_n a_n^2} \right) \right) \\ &= -\log N_n \left(1 - \frac{\rho_n k_n a_n^2}{\log N_n} \times \frac{\nu_{a_n} k_n}{\rho_n k_n a_n^2} \right), \end{aligned}$$

que tende à $-\infty$ quando $n \rightarrow \infty$.

Agora, suponha que $(1 - \rho_n)(N_n)^{2/k_n} \rightarrow \infty$. Observe que,

$$\begin{aligned}
\nu_{a_n} k_n - \log N_n &= \frac{\rho_n k_n a_n^2}{1 + \rho_n} - \frac{k_n \log(1 - \rho_n^2)}{2} - \log N_n \\
&\leq k_n a_n^2 - \frac{k_n \log((1 - \rho_n)(1 + \rho_n))}{2} - \log N_n \\
&= k_n a_n^2 - \frac{k_n \log(1 + \rho_n)}{2} - \frac{k_n \log(1 - \rho_n)}{2} - \log N_n \\
&\leq k_n a_n^2 - \frac{k_n \log(1 - \rho_n)}{2} - \log N_n \\
&= -\frac{k_n}{2} \left(-2a_n^2 + \log(1 - \rho_n) + \frac{2}{k_n} \log N_n \right) \\
&= -\frac{k_n}{2} \left(-2a_n^2 + \log \left((1 - \rho_n) N_n^{\frac{2}{k_n}} \right) \right) \\
&\leq a_n^2 - \frac{1}{2} \log \left((1 - \rho_n) N_n^{\frac{2}{k_n}} \right),
\end{aligned}$$

que tende à $-\infty$ quando a_n cresce mais lentamente que $(1 - \rho_n) N_n^{\frac{2}{k_n}}$. Com isso, concluímos a prova. \square

4.2.2 k_n -conjuntos

Nesta subseção, considere a classe \mathcal{C}_n de todo os conjuntos de tamanho k_n .

Corolário 3 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n a classe de todos os conjuntos de tamanho k_n . Se*

$$\frac{k_n^2}{n} \leq \frac{\log 2}{e^{\nu_1} - 1},$$

então o risco de Bayes satisfaz $R^ \geq 0.3$, e $R^* \rightarrow 1$ se $\frac{k_n^2}{n} \rightarrow \infty$ e $\frac{\rho_n k_n^2}{n} \rightarrow 0$ ou $(1 - \rho_n) n^2 / k_n^4 \rightarrow \infty$ quando $n \rightarrow \infty$.*

Observação 5. Note que a cota inferior de R^* também depende de a_n . Além disso, por um lado note que para $R^* \rightarrow 1$ quando $n \rightarrow \infty$ é preciso que $a_n \rightarrow \infty$. Por outro lado, se $a_n \rightarrow \infty$ muito rápido a função geradora de momentos avaliado em ν_{a_n} pode não ir para 1. Como veremos na prova do Corolário 3 se $\frac{k_n^2}{n} \rightarrow \infty$ e $\frac{\rho_n k_n^2}{n} \rightarrow 0$ ou $(1 - \rho_n) n^2 / k_n^4 \rightarrow \infty$ quando $n \rightarrow \infty$, então é possível encontrar $a_n \rightarrow \infty$ de modo que a função geradora de momentos avaliada em ν_{a_n} converge para 1 fazendo com que o $R^* \rightarrow 1$.

Demonstração. Da Proposição 10, que usa correlação negativa, temos que

$$\mathbb{E} \left[e^{\nu_1 Z} \right] \leq \left((e^{\nu_1} - 1) \frac{k_n}{n} + 1 \right)^{k_n} \leq \exp \left((e^{\nu_1} - 1) \frac{k_n^2}{n} \right),$$

que é limitado por 2 quando $\frac{k_n^2}{n} \leq \frac{\log 2}{e^{\nu_1} - 1}$. Pelo Teorema 6 temos que $R^* \geq 0.3$.

Queremos mostrar que $R^* \rightarrow 1$ quando $n \rightarrow \infty$. Pelo Teorema 6, isso ocorre quando $a_n \rightarrow \infty$ e

$$\exp \left((e^{\nu_1} - 1) \frac{k_n^2}{n} \right) \rightarrow 1.$$

Assim, precisamos mostrar que

$$a_n \rightarrow \infty \quad \text{e} \quad (e^{\nu_{a_n}} - 1) \frac{k_n^2}{n} \rightarrow 0,$$

quando $n \rightarrow \infty$. Suponha que $\frac{k_n^2}{n} \rightarrow \infty$ e $\frac{\rho_n k_n^2}{n} \rightarrow 0$ quando $n \rightarrow \infty$. Desse modo, $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$. Observe que

$$(e^{\nu_{a_n}} - 1) \frac{k_n^2}{n} = \frac{e^{\nu_{a_n}} - 1}{\nu_{a_n}} \times \frac{\nu_{a_n} k_n^2}{n}.$$

Queremos mostrar que $(e^{\nu_{a_n}} - 1)/\nu_{a_n} \rightarrow 1$ e $\nu_{a_n} k_n^2/n \rightarrow 0$ quando $n \rightarrow \infty$. Note que

$$\frac{\nu_{a_n} k_n^2}{n} = \frac{\nu_{a_n}}{\rho_n a_n} \times \frac{\rho_n a_n k_n^2}{n}.$$

Sabemos da prova do Corolário 2 que $\nu_{a_n}/\rho_n a_n^2 \rightarrow 1$ desde que $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$ e além disso, que $a_n \rightarrow \infty$ quando $n \rightarrow \infty$. Basta tomar $a_n \rightarrow \infty$ quando $n \rightarrow \infty$ de modo que $\rho_n a_n k_n^2/n \rightarrow 0$. Como $\frac{k_n^2}{n} \rightarrow \infty$, precisamos que $\rho_n a_n^2 \rightarrow 0$ mais rápido que $\frac{k_n^2}{n} \rightarrow \infty$. Assim, da definição de ν_{a_n} temos que $\nu_{a_n} \rightarrow 0$ e por isso $\frac{e^{\nu_{a_n}} - 1}{\nu_{a_n}} \rightarrow 1$. Segue que $a_n \rightarrow \infty$ e $(e^{\nu_{a_n}} - 1) \frac{k_n^2}{n} \rightarrow 0$.

Agora, suponha que $(1 - \rho_n)n^2/k_n^4 \rightarrow \infty$ quando $n \rightarrow \infty$. Note que

$$(e^{\nu_{a_n}} - 1) \frac{k_n^2}{n} = \frac{e^{\nu_{a_n}} k_n^2}{n} - \frac{k_n^2}{n} \rightarrow 0,$$

quando $n \rightarrow \infty$ desde que $e^{\nu_{a_n}} k_n^2 \ll n$ e $k_n^2 \ll n$. Pela definição de ν_{a_n} é fácil ver que

$$\nu_{a_n} \leq a_n^2 - \frac{1}{2} \log(1 - \rho_n),$$

logo

$$e^{\nu_{a_n}} \leq \frac{e^{a_n^2}}{\sqrt{1 - \rho_n}}.$$

Desse modo, temos que

$$\frac{e^{\nu_{a_n}} k_n^2}{n} \leq \frac{e^{a_n^2} k_n^2}{n\sqrt{1 - \rho_n}} = \exp\left(a_n^2 - \frac{1}{2} \log(1 - \rho_n) + \log(k_n^2/n)\right),$$

que converge à zero se $\log(k_n^2/n) - \frac{1}{2} \log(1 - \rho_n) \rightarrow -\infty$ mais rápido que $a_n^2 \rightarrow \infty$. Observe que

$$\log(k_n^2/n) - \frac{1}{2} \log(1 - \rho_n) = \log\left(\frac{k_n^2}{n\sqrt{1 - \rho_n}}\right),$$

tende à $-\infty$ desde que $\frac{k_n^2}{n\sqrt{1 - \rho_n}} \rightarrow 0$, o que acontece por hipótese. Com isso, concluímos a prova. \square

4.2.3 k_n -intervalos

Nesta subseção, considere a classe \mathcal{C}_n dos k_n -intervalos.

Corolário 4 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n a classe dos k_n -intervalos. Se*

$$\nu_1 \leq \frac{\log\left(\frac{n}{2k_n} - \frac{1}{2}\right)}{k_n},$$

então o risco de Bayes satisfaz $R^ \geq 0.3$, e $R^* \rightarrow 1$ se $\rho_n \ll \min(1, \log(n/k_n)/k_n)$ ou $(1 - \rho_n)(n/k_n)^{2/k_n} \rightarrow \infty$ quando $n \rightarrow \infty$.*

Observação 6. Note que a cota inferior de R^* também depende de a_n . Além disso, por um lado note que para $R^* \rightarrow 1$ quando $n \rightarrow \infty$ é preciso que $a_n \rightarrow \infty$. Por outro lado, se $a_n \rightarrow \infty$ muito rápido a função geradora de momentos avaliado em ν_{a_n} pode não ir para 1. Como veremos na prova do Corolário 4 se $\rho_n \ll \min(1, \log(n/k_n)/k_n)$ ou $(1 - \rho_n)(n/k_n)^{2/k_n} \rightarrow \infty$ quando $n \rightarrow \infty$, então é possível encontrar $a_n \rightarrow \infty$ de modo que a função geradora de momentos avaliada em ν_{a_n} converge para 1 fazendo com que o $R^* \rightarrow 1$.

Demonstração. Seja $[t, t + (k_n - 1)]$ um intervalo de tamanho k_n com $t \geq 1$. Logo, $t + (k_n - 1) \leq n$ implica que $t \leq n - k_n + 1$. Logo, o número de intervalos de tamanho k_n que podemos ter em $[1, \dots, n]$ é $N_n = n - k_n + 1$. Sejam dois intervalos de tamanho k_n escolhidos da seguinte maneira:

$$[t, t + (k_n - 1)] \quad \text{e} \quad [t + (k_n - 1) - (l - 1), t + 2(k_n - 1) - (l - 1)],$$

para $t \geq 1$ e $l = 1, \dots, k_n$. Veja que a interseção entre eles é l e como é necessário que $t + 2(k_n - 1) - (l - 1) \leq n$, logo $t \leq n - 2k_n + l - 1 = N_n - k_n + l$.

Segue que o número de maneiras de se escolher dois k_n -intervalos de modo que a interseção entre eles seja l é $2(N_n - k_n + l)$. Assim, para dois k_n -intervalos escolhidos ao acaso e de maneira independente temos que

$$\mathbb{P}\{|S \cap S'| = l\} = \frac{2(N_n - (k_n - l))}{N_n^2} \quad \text{para todo } l = 1, \dots, k_n.$$

Como $N_n = n - k_n + 1$, logo

$$\begin{aligned} \mathbb{E}\left[e^{\nu Z}\right] - 1 &= \sum_{l=0}^{k_n} e^{\nu l} \mathbb{P}\{|S \cap S'| = l\} - 1 \\ &= \mathbb{P}\{|S \cap S'| = 0\} + \sum_{l=1}^{k_n} e^{\nu l} \mathbb{P}\{|S \cap S'| = l\} - 1 \\ &= \sum_{l=1}^{k_n} (e^{\nu l} - 1) \mathbb{P}\{|S \cap S'| = l\} \\ &\leq \sum_{l=1}^{k_n} e^{\nu l} \mathbb{P}\{|S \cap S'| = l\} \\ &= \sum_{l=1}^{k_n} \frac{2e^{\nu l} (N_n - (k_n - l))}{N_n^2} \\ &\leq \frac{2k_n e^{\nu k_n}}{N_n}, \end{aligned}$$

que é limitado por 1 quando $\nu_1 \leq \frac{\log(N_n/2k_n)}{k_n}$. Desse modo, $\nu_1 \leq \frac{\log\left(\frac{n}{2k_n} - \frac{1}{2}\right)}{k_n}$ e pelo Teorema 6 segue que $R^* \geq 0.3$.

Pelo Teorema 6, para $R^* \rightarrow 1$ quando $n \rightarrow \infty$ é preciso que $a_n \rightarrow \infty$ e $\frac{2k_n e^{\nu a_n k_n}}{N_n} \rightarrow 0$. Logo, preciso mostrar que

$$a_n \rightarrow \infty \quad \text{e} \quad \nu a_n k_n - \log(N_n/k_n) \rightarrow -\infty,$$

quando $n \rightarrow \infty$.

Suponha que $\rho_n \ll \min(1, \log(n/k_n)/k_n)$. Logo $\rho_n \rightarrow 0$ e $\frac{\rho_n k_n}{\log(n/k_n)} \rightarrow 0$ quando $n \rightarrow \infty$. Sabemos da prova do Corolário 2 que $\nu a_n k_n / \rho_n k_n a_n^2 \rightarrow 1$ se $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$, e além disso $a_n \rightarrow \infty$ quando $n \rightarrow \infty$. Escolhemos $a_n \rightarrow \infty$ quando $n \rightarrow \infty$ de modo que $\frac{\rho_n k_n a_n^2}{\log(n/k_n)} \rightarrow 0$.

Como $N_n = n - k_n + 1 = n - (k_n - 1) \leq n$, logo

$$\begin{aligned} \nu_{a_n} k_n - \log(N_n/k_n) &= -\log(N_n/k_n) \left(1 - \frac{\nu_{a_n} k_n}{\log(N_n/k_n)} \right) \\ &= -\log(N_n/k_n) \left(1 - \frac{\nu_{a_n} k_n}{\rho_n k_n a_n^2} \frac{\rho_n k_n a_n^2}{\log(N_n/k_n)} \right) \\ &\leq -\log(N_n/k_n) \left(1 - \frac{\nu_{a_n} k_n}{\rho_n k_n a_n^2} \frac{\rho_n k_n a_n^2}{\log(n/k_n)} \right), \end{aligned}$$

que tende à $-\infty$ quando $n \rightarrow \infty$ pois $N_n/k_n \rightarrow \infty$, $\nu_{a_n} k_n / \rho_n k_n a_n^2 \rightarrow 1$ se $\rho_n \rightarrow 0$ e $\rho_n k_n a_n^2 / \log(n/k_n) \rightarrow 0$.

Agora, suponha que $(1 - \rho_n)(n/k_n)^{2/k_n} \rightarrow \infty$ quando $n \rightarrow \infty$. De maneira análoga a prova feita no Corolário 2 temos que

$$\begin{aligned} \nu_{a_n} k_n - \log(N_n/k_n) &\leq a_n^2 - \frac{1}{2} \log \left((1 - \rho_n)(N_n/k_n)^{\frac{2}{k_n}} \right) \\ &\leq a_n^2 - \frac{1}{2} \log \left((1 - \rho_n)(n/k_n)^{\frac{2}{k_n}} \right), \end{aligned}$$

que tende à $-\infty$ quando $n \rightarrow \infty$ desde que $(1 - \rho_n)(n/k_n)^{2/k_n} \rightarrow \infty$ mais rápido que $a_n^2 \rightarrow \infty$ de modo que $a_n^2 - \frac{1}{2} \log \left((1 - \rho_n)(n/k_n)^{\frac{2}{k_n}} \right) \rightarrow -\infty$. Com isso, concluímos a prova. □

4.2.4 Árvores geradoras

Nesta subseção, considere a classe \mathcal{C}_n de todas as árvores geradoras do grafo completo com $k_n + 1$ vértices. Dessa forma, o número de arestas de grafo é $n = k_n(k_n + 1)/2$.

Corolário 5 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n a classe das árvores geradoras. Existe $\rho^* > 0$ tal que se $\rho_n \leq \rho^*$, então $R^* \geq 0.15$. Além disso, temos que $R^* \rightarrow 1$ quando $n \rightarrow \infty$ desde que $\rho_n \rightarrow 0$.*

Demonstração. Pelo Teorema 6 temos que para $R^* \geq 0.15$ quando tomamos $a = 1$, é preciso que

$$\sqrt{\mathbb{E}[e^{\nu_1 Z}] - 1} \leq \frac{3}{2}.$$

Assim, necessitamos que

$$\mathbb{E}[e^{\nu_1 Z}] \leq \frac{13}{4}.$$

Do Exemplo 3.3.4 temos que

$$\mathbb{E} \left[e^{\nu_1 Z} \right] \leq \left((e^{\nu_1} - 1) \frac{2}{k_n + 1} + 1 \right)^{k_n} \leq e^{2(e^{\nu_1} - 1)}, \quad (4.4)$$

que é limitado por 13/4 quando

$$\nu_1 \leq 1 + \log \left(\log (13/4) / 2 \right).$$

Veja que

$$\begin{aligned} \nu_1 &= \frac{\rho_n}{1 + \rho_n} - \frac{1}{2} \log (1 - \rho_n^2) \\ &= 1 - \frac{1}{1 + \rho_n} - \frac{1}{2} \log (1 - \rho_n^2). \end{aligned}$$

Logo para

$$\nu_1 \leq 1 + \log \left(\log (13/4) / 2 \right),$$

é preciso que

$$-\frac{1}{1 + \rho_n} - \frac{1}{2} \log (1 - \rho_n^2) \leq \log \left(\log (13/4) / 2 \right). \quad (4.5)$$

Como

$$-\frac{1}{1 + \rho_n} - \frac{1}{2} \log (1 - \rho_n^2) \rightarrow -1,$$

quando $\rho_n \rightarrow 0$ e $\log \left(\log (13/4) / 2 \right) > -1$, logo pelo Teorema do Valor Intermediário existe $\rho^* > 0$ tal que (4.5) é satisfeita para todo $0 \leq \rho_n \leq \rho^*$.

Suponha que $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$. Queremos mostrar que $R^* \rightarrow 1$ quando $n \rightarrow \infty$. Veja que por (4.4) e pelo Teorema 6 temos que

$$R^* \geq \mathbb{P}\{|\mathcal{N}(0, 1)| \leq a_n\} \left(1 - \frac{1}{2} \sqrt{e^{2(e^{\nu_{a_n}} - 1)} - 1}\right) \rightarrow 1,$$

quando $n \rightarrow \infty$ desde que $\nu_{a_n} \rightarrow 0$. Sabemos da prova do Corolário 2 que $\nu_{a_n} / \rho_n a_n^2 \rightarrow 1$ desde que $\rho_n \rightarrow 0$ quando $n \rightarrow \infty$ e além disso, que $a_n \rightarrow \infty$ quando $n \rightarrow \infty$. Assim, podemos tomar sequência a_n tal que $a_n \rightarrow \infty$ de modo que $\rho_n a_n^2 \rightarrow 0$. Segue da definição de ν_{a_n} que $\nu_{a_n} \rightarrow 0$ quando $n \rightarrow \infty$. \square

Observação 7. Pode-se mostrar com uma análise mais cuidadosa que a desigualdade em ρ_n é satisfeita para todo $\rho_n \leq 0.4$, isto é, $\rho^* = 0.4$.

4.3 Alguns testes auxiliares

Embora o teste da razão de verossimilhança é ótimo para o risco médio, é preciso estudar testes auxiliares por alguns motivos.

- Primeiro, o teste da razão de verossimilhança é difícil de calcular na maioria das situações.
- Segundo, o teste da razão de verossimilhança é extremamente dependente da priori escolhida, que no nosso caso é a distribuição uniforme na classe \mathcal{C}_n .
- Por fim, é difícil obter diretamente cotas superiores para o risco do pior caso possível do teste da razão de verossimilhança.

Nas subseções a seguir, estudamos o teste do quadrado da soma (TQS), o teste da razão de verossimilhança generalizado (TRVG) e o teste da bondade de ajuste (TBA) e outras variantes. Para cada teste auxiliar, indicamos regimes sob os quais o risco minimax e o risco do pior caso possível para o teste auxiliar têm o mesmo comportamento assintótico. Nesses casos, o teste auxiliar tem o mesmo desempenho do teste ótimo e desse modo temos total entendimento das limitações e possibilidades de detecção do caso considerado.

4.3.1 Teste do Quadrado da Soma

O teste do quadrado da soma é definido em termos da estatística do quadrado da soma $(\sum_{i=1}^n X_i)^2$ que pode ser substancialmente diferente sob \mathbb{P}_0 e \mathbb{P}_S . Mais precisamente veja que, sob \mathbb{P}_0 , a soma $\sum_{i=1}^n X_i$ segue distribuição $\mathcal{N}(0, n)$. Como $\mathcal{N}(0, n)$ segue distribuição $\sqrt{n}\mathcal{N}(0, 1)$, então nesse caso $(\sum_{i=1}^n X_i)^2$ segue distribuição $n\chi_1^2$. Por outro lado, para cada $S \subset \{1, \dots, n\}$ com $|S| = k_n$, note que sob \mathbb{P}_S temos que $\sum_{i=1}^n X_i$ segue distribuição $\mathcal{N}(0, n + \rho_n k_n(k_n - 1))$ e então, de maneira análoga, temos que $(\sum_{i=1}^n X_i)^2$ segue distribuição $(n + \rho_n k_n(k_n - 1))\chi_1^2$.

Proposição 14 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n classe arbitrária de conjuntos de tamanho k_n e suponha que $\rho_n k_n^2/n \rightarrow \infty$ quando $n \rightarrow \infty$. Se $t_n \rightarrow \infty$ quando $n \rightarrow \infty$ de modo que $\frac{t_n}{\rho_n k_n^2/n} \rightarrow 0$, então o risco do pior caso possível do TQS que rejeita a hipótese nula se $(\sum_{i=1}^n X_i)^2 \geq nt_n$, converge para zero quando $n \rightarrow \infty$. Por outro lado, para todo teste baseado na soma $(\sum_{i=1}^n X_i)^2$ temos que o risco do pior caso possível converge para 1 se $\rho_n k_n^2/n \rightarrow 0$.*

Demonstração. Defina o teste do quadrado da soma como sendo

$$f_{TQS}(X) = \begin{cases} 1, & \text{se } \left(\sum_{i=1}^n X_i\right)^2 \geq nt_n, \\ 0, & \text{caso contrário.} \end{cases}$$

Suponha que $\rho_n k_n^2/n \rightarrow \infty$ e $\frac{t_n}{\rho_n k_n^2/n} \rightarrow 0$. Desse modo,

$$\begin{aligned} R_{\max}(f_{TQS}) &= \mathbb{P}_0 \left\{ \left(\sum_{i=1}^n X_i \right)^2 \geq nt_n \right\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \left(\sum_{i=1}^n X_i \right)^2 < nt_n \right\} \\ &= \mathbb{P} \left\{ n\chi_1^2 \geq nt_n \right\} + \mathbb{P} \left\{ (n + \rho_n k_n(k_n - 1))\chi_1^2 < nt_n \right\} \\ &\leq \frac{1}{t_n} + \mathbb{P} \left\{ \chi_1^2 < \frac{1}{\frac{1}{t_n} + \frac{\rho_n k_n(k_n - 1)}{nt_n}} \right\}, \end{aligned}$$

que converge para zero quando $n \rightarrow \infty$.

Agora, suponha que $\rho_n k_n^2/n \rightarrow 0$. Desse modo,

$$\begin{aligned} R_{\max}(f_{TQS}) &= \mathbb{P}_0 \left\{ \left(\sum_{i=1}^n X_i \right)^2 \geq t_n \right\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \left(\sum_{i=1}^n X_i \right)^2 < t_n \right\} \\ &= \mathbb{P} \left\{ n\chi_1^2 \geq t_n \right\} + \mathbb{P} \left\{ (n + \rho_n k_n(k_n - 1))\chi_1^2 < t_n \right\} \\ &= \mathbb{P} \left\{ \chi_1^2 \geq \frac{t_n}{n} \right\} + \mathbb{P} \left\{ \chi_1^2 < \frac{t_n/n}{1 + \frac{\rho_n k_n(k_n - 1)}{n}} \right\} \\ &= 1 - \mathbb{P} \left\{ t_n/n \geq \chi_1^2 \geq \frac{t_n/n}{1 + \frac{\rho_n k_n(k_n - 1)}{n}} \right\}, \end{aligned}$$

que converge para um quando $n \rightarrow \infty$ pois $\rho_n k_n^2/n \rightarrow 0$ por hipótese. \square

No caso dos k_n -conjuntos e no regime $k_n^2/n \rightarrow \infty$ temos que:

- Se $k_n^2/n \ll 1/\rho_n$, então pelo Corolário 3 temos que $R_{\max}^* \geq R^* \rightarrow 1$ e segue que $R_{\max}(f_{TQS}) \rightarrow 1$.
- Se $k_n^2/n \gg 1/\rho_n$, então pela Proposição 14 temos que $R_{\max}(f_{TQS}) \rightarrow 0$. Como $R_{\max}^* \leq R_{\max}(f_{TQS}) \rightarrow 0$, segue que $R_{\max}^* \rightarrow 0$.

Portanto, o TQS é tão bom quanto o teste ótimo (teste da razão de verossimilhança) para detectar a presença de variáveis anômalas na classe dos k_n -conjuntos nos casos $k_n^2/n \ll 1/\rho_n$ e $k_n^2/n \gg 1/\rho_n$, desde que $k_n^2/n \rightarrow \infty$.

Por outro lado, no regime $k_n^2/n \rightarrow 0$, temos pela Proposição 14 que $R_{\max}(f_{TQS}) \rightarrow 1$ para todo f_{TQS} . Assim, nesse regime, o TQS não nos permite detectar a presença de variáveis anômalas na classe dos k_n -conjuntos.

4.3.2 Teste da Razão de Verossimilhança Generalizado

Nesta subseção, investigamos a performance do teste da razão de verossimilhança generalizado (TRVG) na classe dos k_n -intervalos e na classe dos k_n -conjuntos. Na classe dos k_n -intervalos, o TRVG tem o mesmo desempenho do teste ótimo. No entanto, na classe dos k_n -conjuntos, o TRVG se comporta mal em alguns regimes.

Por definição, o TRVG rejeita a hipótese nula H_0 , definida em (2.2), para grandes valores de $\max_{S \in \mathcal{C}_n} \frac{Z_S}{\sqrt{\det(A_S)}}$, onde Z_S é definido em (2.10) e A_S é definido em (2.3). Como $\det(A_S) = (1 - \rho_n)^{k_n}(1 + \rho_n(k_n - 1))$ não depende de S , logo o TRVG é da forma

$$f_{TRVG}(X) = 0 \text{ se e somente se } \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \leq t_n,$$

para algum t_n escolhido de forma apropriada. Notamos imediatamente que o TRVG requer conhecimento de ρ_n .

A nossa análise é baseada no Lema 9 que fornece a distribuição da variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ sob \mathbb{P}_0 e sob \mathbb{P}_S . Sob \mathbb{P}_0 , precisamos controlar o máximo da variável aleatória $X^T(\mathbf{I}_n - A_S^{-1})X$ sobre $S \in \mathcal{C}_n$, que é feito usando desigualdades de concentração para a distribuição qui-quadrado.

TRVG para k_n -intervalos. Pelo Corolário 4, o risco de Bayes tende à 1 quando $\rho_n \rightarrow 0$ e $\frac{\rho_n k_n}{\log(n/k_n)} \rightarrow 0$. Suponha que $k_n/\log n \rightarrow \infty$. Vamos enunciar o seguinte resultado que é aplicado a classes com cardinalidade pequena.

Proposição 15 (Arias-Castro et al. (2012)). *Considere a classe \mathcal{C}_n dos conjuntos de tamanho k_n , com cardinalidade $N_n \rightarrow \infty$ tal que $(\log N_n)/k_n \rightarrow 0$. Quando $\rho_n k_n/\log N_n \rightarrow \infty$, o risco do pior caso possível do TRVG com $t_n = -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} + e \log N_n$ converge à zero.*

Demonstração. Veja que para o TRVG,

$$R_{\max}(f_{TRVG}) = \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \geq t_n \right\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \leq t_n \right\}.$$

Primeiramente, vamos dar cota superior para a probabilidade

$$\mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \geq t_n \right\}.$$

Veja que, sob \mathbb{P}_0 , pelo Lema 9 temos que

$$X^T(\mathbf{I}_n - A_S^{-1})X = \frac{-\rho_n}{1 - \rho_n}C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}D_S,$$

onde $C_S \sim \chi_{k_n-1}^2$ e $D_S \sim \chi_1^2$. Logo,

$$\begin{aligned} \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X &= \max_{S \in \mathcal{C}_n} \left\{ \frac{-\rho_n}{1 - \rho_n}C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}D_S \right\} \\ &= \frac{-\rho_n}{1 - \rho_n}C_{S_0} + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}D_{S_0} \\ &\leq -\rho_n C_{S_0} + D_{S_0} \\ &\leq -\rho_n \min_{S \in \mathcal{C}_n} C_S + \max_{S \in \mathcal{C}_n} D_S, \end{aligned} \quad (4.6)$$

onde S_0 é o conjunto onde $\frac{\rho_n}{1 - \rho_n}C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)}D_S$ é maximizado. Seja $t_n = \alpha_n + \beta_n$. Logo,

$$\begin{aligned} \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \geq t_n \right\} &= \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I} - A_S^{-1})X \geq \alpha_n + \beta_n \right\} \\ &\stackrel{(i)}{\leq} \mathbb{P} \left\{ -\rho_n \min_{S \in \mathcal{C}_n} C_S + \max_{S \in \mathcal{C}_n} D_S \geq \alpha_n + \beta_n \right\} \\ &\stackrel{(ii)}{\leq} \mathbb{P} \left\{ -\rho_n \min_{S \in \mathcal{C}_n} C_S \geq \alpha_n \right\} + \mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq \beta_n \right\}, \end{aligned} \quad (4.7)$$

onde em (i) usamos (4.6) e em (ii) usamos a cota da união. Como $D_S \sim \chi_1^2$, logo

$$\mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq \beta_n \right\} = \mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq \sqrt{\beta_n} \right\},$$

onde $Z_i \sim \mathcal{N}(0, 1)$ independentes para todo $i = 1, \dots, N_n$. O seguinte resultado fornece cota superior para a igualdade acima.

Lema 10. *Sejam as variáveis aleatórias $Z_i \sim \mathcal{N}(0, 1)$ independentes para todo $i = 1, \dots, N_n$. Desse modo,*

$$\mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq x \right\} \leq \exp(\log N_n - x^2/2).$$

Demonstração. Pelo truque de Chernoff temos que para $\lambda > 0$,

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq x \right\} &\leq e^{-\lambda x} \mathbb{E} \left[e^{\lambda \max_{1 \leq i \leq N_n} Z_i} \right] \\ &\leq e^{-\lambda x} \sum_{i=1}^{N_n} \mathbb{E} \left[e^{\lambda Z_i} \right] \\ &= e^{-\lambda x} N_n e^{\lambda^2/2} \\ &= \exp \left(-\lambda x + \log N_n + \lambda^2/2 \right). \end{aligned}$$

Minimizamos em $\lambda > 0$ a função $\exp \left(-\lambda x + \log N_n + \lambda^2/2 \right)$ temos que

$$\mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq x \right\} \leq \exp \left(\log N_n - x^2/2 \right).$$

□

Assim, pelo Lema 10 concluímos que

$$\mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq \beta_n \right\} = \mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq \sqrt{\beta_n} \right\} \leq \exp \left(\log N_n - \beta_n/2 \right).$$

Tomando $\beta_n = e \log N_n$ temos que

$$\mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq \beta_n \right\} = \mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq e \log N_n \right\} \leq \exp \left(\log N_n - \frac{e}{2} \log N_n \right), \quad (4.8)$$

que converge à zero quando $n \rightarrow \infty$. Agora, note que

$$\begin{aligned} \mathbb{P} \left\{ -\rho_n \min_{S \in \mathcal{C}_n} C_S \geq \alpha_n \right\} &= \mathbb{P} \left\{ \min_{S \in \mathcal{C}_n} C_S \leq \frac{-\alpha_n}{\rho_n} \right\} \\ &\stackrel{(iii)}{\leq} N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq \frac{-\alpha_n}{\rho_n} \right\}, \end{aligned}$$

onde em (iii) usamos a cota da união. Queremos mostrar para um certo α_n , a expressão $N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq \frac{-\alpha_n}{\rho_n} \right\} \rightarrow 0$ quando $n \rightarrow \infty$. Tomando $\frac{-\alpha_n}{\rho_n} = a(k_n - 1)$ para $a \in (0, 1)$ e usando o truque de Chernoff temos que

$$\begin{aligned} \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq a(k_n - 1) \right\} &\leq e^{\lambda a(k_n-1)} \mathbb{E} \left[e^{-\lambda \chi_{k_n-1}^2} \right] \\ &\stackrel{(iv)}{=} e^{\lambda a(k_n-1)} (1 + 2\lambda)^{-(k_n-1)/2} \\ &= \exp \left(- \left(\frac{k_n - 1}{2} \right) (\log(1 + 2\lambda) - 2\lambda a) \right), \end{aligned} \quad (4.9)$$

onde em (iv) usamos a função geradora de momentos da variável aleatória $\chi_{k_n-1}^2$. Desse modo, minimizando em $\lambda > 0$ a última igualdade acima encontramos $\lambda = \frac{1-a}{2a}$ e com isso temos que

$$\mathbb{P} \left\{ \chi_{k_n-1}^2 \leq a(k_n - 1) \right\} \leq \exp \left(- \left(\frac{k_n - 1}{2} \right) (-\log a + a - 1) \right).$$

Veja que quando $a \rightarrow 1$, pela expansão de Taylor

$$\log a \approx 1 - a + \frac{(1-a)^2}{2}.$$

Desse modo, quando $a \rightarrow 1$ temos que

$$-\log a + a - 1 \approx \frac{(1-a)^2}{2}.$$

Assim, tomando $a = 1 - \sqrt{(5 \log N_n)/k_n}$, multiplicando por N_n em ambos os lados de (4.9) e usando a aproximação acima (note que $a \rightarrow 1$ pois $\log N_n/k_n \rightarrow 0$ por hipótese) temos que,

$$\begin{aligned} \limsup_{n \rightarrow \infty} N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq (1 - \sqrt{(5 \log N_n)/k_n})(k_n - 1) \right\} \\ \leq \limsup_{n \rightarrow \infty} \exp \left(- \frac{5(k_n - 1)}{4k_n} \log N_n + \log N_n \right) \\ = \limsup_{n \rightarrow \infty} \exp \left(- \frac{\log N_n}{4} + \frac{5 \log N_n}{4k_n} \right) \stackrel{(v)}{=} 0, \quad (4.10) \end{aligned}$$

onde em (v) usamos que $N_n \rightarrow \infty$ e $\log N_n/k_n \rightarrow 0$ quando $n \rightarrow \infty$. A partir da definição de α_n concluímos que

$$\alpha_n = -\rho_n a(k_n - 1) = -\rho_n(k_n - 1) + \rho_n \sqrt{\frac{5(k_n - 1)^2}{k_n} \log N_n}.$$

Como quando $n \rightarrow \infty$, k_n e $k_n - 1$ têm a mesma ordem, temos

$$\alpha_n \approx -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} = \tilde{\alpha}_n.$$

Assim, combinando (4.8) e (4.10) em (4.7) temos que a probabilidade

$$\mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \geq t_n \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$ com $t_n = \tilde{\alpha}_n + e \log N_n$. Agora, vamos dar cota superior para a probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\}.$$

Note que, sob \mathbb{P}_S , pelo Lema 9 temos que

$$\begin{aligned} \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\} &\leq \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\} \\ &= \max_{S \in \mathcal{C}_n} \mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 + \rho_n (k_n - 1) \chi_1^2 < t_n \right\} \\ &= \mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 + \rho_n (k_n - 1) \chi_1^2 \leq t_n \right\}. \end{aligned}$$

Tomando $t_n = -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} + e \log N_n$ e usando a cota da união temos que a probabilidade acima é limitada superiormente por

$$\mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 \leq -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} \right\} + \mathbb{P} \left\{ \rho_n (k_n - 1) \chi_1^2 \leq e \log N_n \right\}.$$

Veja que

$$\mathbb{P} \left\{ \rho_n (k_n - 1) \chi_1^2 \leq e \log N_n \right\} = \mathbb{P} \left\{ \chi_1^2 \leq e \log N_n / \rho_n (k_n - 1) \right\},$$

que converge à zero pois $\rho_n k_n / \log N_n \rightarrow \infty$ quando $n \rightarrow \infty$. Note que,

$$\begin{aligned} \mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 \leq -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} \right\} &= \mathbb{P} \left\{ \chi_{k_n-1}^2 \geq k_n - \sqrt{5k_n \log N_n} \right\} \\ &= \mathbb{P} \left\{ \chi_{k_n-1}^2 - (k_n - 1) \right. \\ &\quad \left. \geq 1 - \sqrt{5k_n \log N_n} \right\} \\ &\leq \mathbb{P} \left\{ \frac{|\chi_{k_n-1}^2 - (k_n - 1)|}{\sqrt{2(k_n - 1)}} \right. \\ &\quad \left. \geq \frac{\sqrt{5k_n \log N_n} - 1}{\sqrt{2(k_n - 1)}} \right\} \\ &\stackrel{(vi)}{\leq} \frac{2(k_n - 1)}{(\sqrt{5k_n \log N_n} - 1)^2} \\ &= \frac{2}{\left(\sqrt{\frac{5k_n \log N_n}{k_n - 1}} - \frac{1}{\sqrt{k_n - 1}} \right)^2}, \end{aligned}$$

onde em (vi) usamos a desigualdade de Markov. Observe que a probabilidade acima converge à zero quando $n \rightarrow \infty$ pois como

$$\frac{5k_n \log N_n}{k_n - 1} \rightarrow \infty \text{ e } \frac{1}{\sqrt{k_n - 1}} \rightarrow 0,$$

quando $n \rightarrow \infty$ temos que $\left(\sqrt{\frac{5k_n \log N_n}{k_n-1}} - \frac{1}{\sqrt{k_n-1}}\right)^2 \rightarrow \infty$. Portanto, a probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$ com $t_n = -\rho_n k_n + \rho_n \sqrt{5k_n \log N_n} + e \log N_n$. Segue que o risco do pior caso possível converge à zero. \square

No caso dos k_n -intervalos temos que $N_n = n - k_n + 1$ e desse modo,

$$\begin{aligned} \frac{\log N_n}{k_n} &= \frac{1}{k_n} \log \left(\frac{n}{k_n} \left(k_n - \frac{k_n^2}{n} + \frac{k_n}{n} \right) \right) \\ &= \frac{\log(n/k_n)}{k_n} + \frac{1}{k_n} \log \left(k_n - \frac{k_n^2}{n} + \frac{k_n}{n} \right). \end{aligned}$$

Observe que

$$\frac{1}{k_n} \log \left(k_n - \frac{k_n^2}{n} + \frac{k_n}{n} \right) = \frac{\log k_n}{k_n} + \frac{1}{k_n} \log \left(1 - \frac{k_n}{n} + \frac{1}{n} \right),$$

que converge à zero quando $n \rightarrow \infty$. Logo para $(\log N_n)/k_n \rightarrow 0$, basta verificarmos se $\log(n/k_n)/k_n \rightarrow 0$, ou seja, se $(\log n)/k_n \rightarrow 0$. Veja que a hipótese $\rho_n \ll \min \left\{ 1, \frac{1}{k_n} \log \left(\frac{n}{k_n} \right) \right\}$ presente no Corolário 2 é equivalente

à

$$\rho_n \ll \min \left\{ 1, \frac{1}{k_n} \log \left(\frac{n}{k_n} \right) \right\} = \min \left\{ 1, \frac{\log n - \log k_n}{k_n} \right\} = \frac{\log n}{k_n},$$

se $k_n/\log n \rightarrow \infty$.

Desse modo, no caso dos k_n -intervalos e no regime $k_n/\log n \rightarrow \infty$ temos que:

- Se $k_n/\log n \ll 1/\rho_n$, então pelo Corolário 4 temos que $R^* \rightarrow 1$. Como $R_{\max}(f_{TRVG}) \geq R_{\max}^* \geq R^* \rightarrow 1$, segue que $R_{\max}(f_{TRVG}) \rightarrow 1$.
- Se $k_n/\log n \gg 1/\rho_n$, então pela Proposição 15 temos que $R_{\max}(f_{TRVG}) \rightarrow 0$. Como $R_{\max}^* \leq R_{\max}(f_{TRVG}) \rightarrow 0$, segue que $R_{\max}^* \rightarrow 0$.

Portanto, o TRVG é tão bom quanto o teste ótimo (teste da razão de verossimilhança) para detectar a presença de variáveis anômalas na classe dos k_n -intervalos nos casos $k_n/\log n \ll 1/\rho_n$ e $k_n/\log n \gg 1/\rho_n$, desde que $k_n/\log n \rightarrow \infty$.

TRVG para k_n -conjuntos. Considere agora o exemplo da classe de todos os k_n -conjuntos. A situação aqui é diferente pois N_n , o tamanho da classe \mathcal{C}_n , é muito maior. Por exemplo, no caso dos k_n -conjuntos, $N_n = \binom{n}{k_n}$, e portanto $(\log N_n)/k_n \rightarrow \infty$ quando $n \rightarrow \infty$. O equivalente da Proposição 15 para esse regime é o seguinte.

Proposição 16 (Arias-Castro et al. (2012)). *Considere a classe dos conjuntos \mathcal{C}_n dos conjuntos de tamanho k_n , com cardinalidade $N_n \rightarrow \infty$ tal que $(\log N_n)/k_n \rightarrow \infty$. Quando $\eta = (1 - \rho_n)N_n^{2/k_n}(\log N_n)/k_n \rightarrow 0$, o risco do pior caso possível do TRVG com $t_n = -\log N_n \left(\frac{\rho_n}{\sqrt{\eta}} - e \right)$ converge à zero.*

Demonstração. Veja que para o TRVG,

$$R_{\max}(f_{TRVG}) = \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \geq t_n \right\} \\ + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\}.$$

Primeiramente, vamos dar cota superior para a probabilidade

$$\mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \geq t_n \right\}.$$

Veja que, sob \mathbb{P}_0 , pelo Lema 9 temos que

$$X^T (\mathbf{I}_n - A_S^{-1}) X = -\frac{\rho_n}{1 - \rho_n} C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)} D_S,$$

onde $C_S \sim \chi_{k_n-1}^2$ e $D_S \sim \chi_1^2$. De maneira similar a feita na prova da Proposição 15 temos que

$$\max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X = \max_{S \in \mathcal{C}_n} \left\{ \frac{-\rho_n}{1 - \rho_n} C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)} D_S \right\} \\ = \frac{-\rho_n}{1 - \rho_n} C_{S_0} + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)} D_{S_0} \\ \leq \frac{-\rho_n}{1 - \rho_n} C_{S_0} + D_{S_0} \\ \leq \frac{-\rho_n}{1 - \rho_n} \min_{S \in \mathcal{C}_n} C_S + \max_{S \in \mathcal{C}_n} D_S,$$

onde S_0 é o conjunto onde $\frac{\rho_n}{1 - \rho_n} C_S + \frac{\rho_n(k_n - 1)}{1 + \rho_n(k_n - 1)} D_S$ é maximizado.

Seja $t_n = \alpha_n + \beta_n$. Note que,

$$\begin{aligned} \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \geq t_n \right\} &= \mathbb{P}_0 \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \geq \alpha_n + \beta_n \right\} \\ &\leq \mathbb{P} \left\{ -\frac{\rho_n}{1 - \rho_n} \min_{S \in \mathcal{C}_n} C_S + \max_{S \in \mathcal{C}_n} D_S \geq \alpha_n + \beta_n \right\} \\ &\stackrel{(i)}{\leq} \mathbb{P} \left\{ -\frac{\rho_n}{1 - \rho_n} \min_{S \in \mathcal{C}_n} C_S \geq \alpha_n \right\} \\ &\quad + \mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq \beta_n \right\}, \end{aligned}$$

onde em (i) usamos a cota da união. De (4.8) temos que para $\beta_n = e \log N_n$ vale que

$$\mathbb{P} \left\{ \max_{S \in \mathcal{C}_n} D_S \geq e \log N_n \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$. Note que pela cota da união temos que

$$\begin{aligned} \mathbb{P} \left\{ -\frac{\rho_n}{1 - \rho_n} \min_{S \in \mathcal{C}_n} C_S \geq \alpha_n \right\} &= \mathbb{P} \left\{ \min_{S \in \mathcal{C}_n} C_S \leq -\frac{(1 - \rho_n)\alpha_n}{\rho_n} \right\} \\ &\leq N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq -\frac{(1 - \rho_n)\alpha_n}{\rho_n} \right\}. \end{aligned}$$

Tomando $\frac{-(1 - \rho_n)\alpha_n}{\rho_n} = a(k_n - 1)$ para $a \in (0, 1)$ e pela prova da Proposição 15 temos que

$$N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq a(k_n - 1) \right\} \leq \exp \left(- \left(\frac{k_n - 1}{2} \right) \left(-\log a + a - 1 - \frac{2 \log N_n}{k_n - 1} \right) \right).$$

Para que $N_n \mathbb{P} \left\{ \chi_{k_n-1}^2 \leq -\frac{(1 - \rho_n)\alpha_n}{\rho_n} \right\} \rightarrow 0$ quando $n \rightarrow \infty$ e $a \rightarrow 0$ é preciso que

$$\log a + \frac{2 \log N_n}{k_n} \rightarrow -\infty,$$

quando $n \rightarrow \infty$, o que ocorre se e somente se,

$$\log a N_n^{2/k_n} \rightarrow -\infty.$$

Tome $a N_n^{2/k_n} = \sqrt{\eta}$, onde $\eta = (1 - \rho_n) N_n^{2/k_n} (\log N_n) / k_n$. Assim, precisamos que $\rho_n \rightarrow 1$ quando $n \rightarrow \infty$ de modo que $(1 - \rho_n) \rightarrow 0$ mais rápido que $N_n^{2/k_n} (\log N_n) / k_n \rightarrow \infty$. Com essa escolha, $\eta \rightarrow 0$. Substituindo $a = \frac{\sqrt{\eta}}{N_n^{2/k_n}}$

na definição de α_n encontramos $\alpha_n = -\frac{\rho_n}{1-\rho_n}\sqrt{\eta}N_n^{-2/k_n}(k_n-1)$. Desse modo, basta tomarmos

$$\begin{aligned} t_n &= \alpha_n + \beta_n \\ &= -\frac{\rho_n}{1-\rho_n}N_n^{-2/k_n}\sqrt{\eta}k_n + e \log N_n \\ &= \frac{-\rho_n\sqrt{\eta}\log N_n}{(1-\rho_n)N_n^{2/k_n}(\log N_n)/k_n} + e \log N_n \\ &= \frac{-\rho_n \log N_n}{\sqrt{\eta}} + e \log N_n \\ &= -\log N_n \left(\frac{\rho_n}{\sqrt{\eta}} - e \right), \end{aligned}$$

para que a probabilidade $\mathbb{P}_0 \{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \geq t_n \} \rightarrow 0$, quando $n \rightarrow \infty$. Agora, vamos dar cota superior para a probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T(\mathbf{I}_n - A_S^{-1})X \leq t_n \right\}.$$

Veja que, sob \mathbb{P}_S , pelo Lema 9 temos que

$$\begin{aligned} \max_{S \in \mathcal{C}} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}} X^T(\mathbf{I}_n - A_S^{-1})X \leq t_n \right\} &\leq \max_{S \in \mathcal{C}} \mathbb{P}_S \left\{ X^T(\mathbf{I}_n - A_S^{-1})X \leq t_n \right\} \\ &= \max_{S \in \mathcal{C}} \mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 + \rho_n(k_n-1)\chi_1^2 < t_n \right\} \\ &= \mathbb{P} \left\{ -\rho_n \chi_{k_n-1}^2 + \rho_n(k_n-1)\chi_1^2 \leq t_n \right\}. \end{aligned}$$

Tomando $t_n = -\log N_n \left(\frac{\rho_n}{\sqrt{\eta}} - e \right)$ e usando a cota da união temos que a probabilidade acima é limitada superiormente por

$$\mathbb{P} \left\{ \chi_{k_n-1}^2 \geq \frac{\log N_n}{\sqrt{\eta}} \right\} + \mathbb{P} \left\{ \chi_1^2 \geq \frac{e \log N_n}{\rho_n(k_n-1)} \right\}.$$

Como $\rho_n \rightarrow 1$ e $(\log N_n)/k_n \rightarrow \infty$, logo

$$\mathbb{P} \left\{ \chi_1^2 \geq \frac{e \log N_n}{\rho_n(k_n-1)} \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$. Veja que pela desigualdade de Markov,

$$\begin{aligned} \mathbb{P} \left\{ \chi_{k_n-1}^2 \geq \frac{\log N_n}{\sqrt{\eta}} \right\} &\leq \frac{\sqrt{\eta}}{\log N_n}(k_n-1) \\ &= \sqrt{\eta} \left(\frac{k_n-1}{\log N_n} \right), \end{aligned}$$

que converge à zero quando $n \rightarrow \infty$ pois $\eta \rightarrow 0$ e $\log(N_n)/k_n \rightarrow \infty$. Portanto, a probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S \in \mathcal{C}_n} X^T (\mathbf{I}_n - A_S^{-1}) X \leq t_n \right\} \rightarrow 0,$$

e assim concluímos que o risco do pior caso possível converge à zero. \square

Primeiro, observe que para os k_n -conjuntos, $N_n = \binom{n}{k_n}$ e como $(n/k_n)^{k_n} \leq N_n \leq (ne/k_n)^{k_n}$, temos que $k_n \log(n/k_n) \leq \log N_n \leq k_n \log(n/k_n) + k_n$ e $(n/k_n)^2 \leq N_n^{2/k_n} \leq (ne/k_n)^2$. Assim, $(1 - \rho_n) N_n^{2/k_n} (\log N_n)/k_n$ é da ordem de $(1 - \rho_n)(n/k_n)^2 \log(n/k_n)$.

Pela Proposição 16 temos que se $\log(n/k_n) \rightarrow \infty$ e $(1 - \rho_n)(n/k_n)^2 \log(n/k_n) \rightarrow 0$, então $R_{\max}(f_{TRVG}) \rightarrow 0$. Observe que pelas Proposições 16 e 14, respectivamente, temos que:

- Se $\log(n/k_n) \rightarrow \infty$ e $(1 - \rho_n)(n/k_n)^2 \log(n/k_n) \rightarrow 0$, então $R_{\max}(f_{TRVG}) \rightarrow 0$.
- Se $k_n^2/n \rightarrow \infty$ e $k_n^2/n \gg 1/\rho_n$, então $R_{\max}(f_{TQS}) \rightarrow 0$.

Do primeiro item temos que $\rho_n \rightarrow 1$ e $n/k_n \rightarrow \infty$ e essa condição é mais forte que a condição requerida no segundo item pois $\rho_n k_n^2/n \rightarrow \infty$ pode valer mesmo que $\rho_n \not\rightarrow 1$ quando $n \rightarrow \infty$. Para ilustrar isso, enunciaremos o seguinte teorema.

Teorema 7. *Seja \mathcal{C}_n a classe dos k_n -conjuntos. Se $\rho_n < 0.6$ e $k_n \ll n^{0.7}$, então o TRVG tem risco médio estritamente positivo.*

A prova do teorema acima não será apresentada nessa dissertação. A sua demonstração pode ser encontrada em (Arias-Castro et al. (2012), ver Teorema 3.1).

Com as escolhas $\rho_n = 0.5$ e $k_n = n^{0.6}$, então pelo Teorema 7 temos que o risco médio para o TRVG é estritamente positivo. Porém, sob o mesmo regime, temos que a condição $\rho_n k_n^2/n \rightarrow \infty$ é satisfeita e pela Proposição 14 temos que o risco do pior caso possível para o TQS converge à zero. Portanto, no regime $k_n^2/n \rightarrow \infty$, o TRVG é “pior” do que TQS, que tem o mesmo desempenho do teste ótimo nesse regime.

Veja que no regime $k_n^2/n \rightarrow 0$, o TQS não é capaz de detectar a presença de variáveis anômalas mesmo que $\rho_n = 1$. Ainda nesse regime, pela Proposição 3 temos que se $1/(1 - \rho_n) \ll n^2/k_n^4$, então $R_{\max}^* \geq R^* \rightarrow 1$. Além disso, pela Proposição 16 temos que se $1/(1 - \rho_n) \gg (n/k_n^2) \log(n/k_n)$, então $R_{\max}^* \leq R_{\max}(f_{TRVG}) \rightarrow 0$. Observe que $(n/k_n^2) \log(n/k_n) \gg n^2/k_n^4$ quando

$k_n^2/n \rightarrow 0$. Desse modo, com base nos resultados obtidos, não temos como determinar no regime $k_n^2/n \rightarrow 0$ o que acontece com o risco do pior caso possível para o TRVG quando $1/(1 - \rho_n) \gg n^2/k_n^4$. Em particular, nesse mesmo regime, com os testes auxiliares utilizados até agora não podemos determinar o comportamento do risco minimax. Essa questão será discutida novamente na subseção 4.3.4.

4.3.3 Teste do Quadrado da Soma localizada

Enquanto que o TRVG é tão bom quanto o teste ótimo para detectar k_n -intervalos, é preciso conhecimento de ρ_n . No entanto, uma simples modificação resolve esse problema. Considere o seguinte teste do quadrado da soma localizada (TQSL):

$$f_{TQSL}(X) = \begin{cases} 0, & \text{se } \max_{S \in \mathcal{C}_n} \left(\sum_{i \in S} X_i \right)^2 \leq t_n, \\ 1, & \text{caso contrário,} \end{cases}$$

para algum t_n apropriado.

Proposição 17 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n uma classe qualquer de conjuntos de tamanho k_n , com cardinalidade $N_n \rightarrow \infty$ tal que $(\log N_n)/k_n \rightarrow 0$. Quando $\rho_n \gg (\log N_n)/k_n$, o risco do pior caso possível para o TQSL com $t_n = ek_n \log N_n$ converge à zero.*

Demonstração. Veja que,

$$R_{\max}(f_{TQSL}) = \mathbb{P}_0 \left\{ \max_{i \in S} \left(\sum_{i \in S} X_i \right)^2 \geq t_n \right\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{i \in S'} \left(\sum_{i \in S'} X_i \right)^2 \leq t_n \right\}.$$

Como sob \mathbb{P}_0 , a variável aleatória $\sum_{i \in S} X_i \sim \mathcal{N}(0, k_n)$, logo temos que

$$\begin{aligned} \mathbb{P}_0 \left\{ \max_{i \in S} \left(\sum_{i \in S} X_i \right)^2 \geq t_n \right\} &= \mathbb{P} \left\{ \max_{i \in S} k_n \chi_1^2 \geq t_n \right\} \\ &= \mathbb{P} \left\{ \max_{1 \leq i \leq N_n} Z_i \geq \sqrt{\frac{t_n}{k_n}} \right\} \\ &\leq \exp(\log N_n - t_n/2k_n), \end{aligned}$$

onde $Z_i \sim \mathcal{N}(0, 1)$ independente para todo $i = 1, \dots, N_n$. Tomando $t_n = ek_n \log N_n$ temos que

$$\mathbb{P}_0 \left\{ \max_{i \in S} \left(\sum_{i \in S} X_i \right)^2 \geq t_n \right\} \leq \exp \left(\log N_n - \frac{e}{2} \log N_n \right),$$

que converge à zero quando $n \rightarrow \infty$.

Agora sob \mathbb{P}_S , a variável aleatória $(\sum_{i \in S} X_i)^2 \sim (k_n + k_n(k_n - 1)\rho_n)\chi_1^2$ e novamente tomando $t_n = ek_n \log N_n$ temos que

$$\begin{aligned} \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{S' \in \mathcal{C}_n} \left(\sum_{i \in S'} X_i \right)^2 \leq ek_n \log N_n \right\} &\leq \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \left(\sum_{i \in S'} X_i \right)^2 \leq ek_n \log N_n \right\} \\ &= \max_{S \in \mathcal{C}_n} \mathbb{P} \left\{ (k_n + k_n(k_n - 1)\rho_n)\chi_1^2 \leq ek_n \log N_n \right\} \\ &= \mathbb{P} \left\{ \chi_1^2 \leq \frac{ek_n \log N_n}{(k_n + k_n(k_n - 1)\rho_n)} \right\} \\ &= \mathbb{P} \left\{ \chi_1^2 \leq \frac{e}{\frac{1}{\log N_n} + \frac{(k_n - 1)\rho_n}{\log N_n}} \right\}, \end{aligned}$$

que converge à zero pois $\rho_n k_n / \log N_n \rightarrow \infty$. Portanto, segue que o risco do pior caso possível tende à zero. \square

Lembre que no caso dos k_n -intervalos, para $(\log N_n)/k_n \rightarrow 0$ quando $n \rightarrow \infty$ basta $(\log n)/k_n \rightarrow 0$. Desse modo, no caso dos k_n -intervalos e no regime $(\log N_n)/k_n \rightarrow 0$ temos que:

- Se $k_n / \log n \ll 1/\rho_n$, então pelo Corolário 4 temos que $R^* \rightarrow 1$. Como $R_{\max}(f_{TQSL}) \geq R_{\max}^* \geq R^* \rightarrow 1$, segue que $R_{\max}(f_{TQSL}) \rightarrow 1$.
- Se $k_n / \log n \gg 1/\rho_n$, então pela Proposição 17 temos que $R_{\max}(f_{TQSL}) \rightarrow 0$. Como $R_{\max}^* \leq R_{\max}(f_{TQSL}) \rightarrow 0$, segue que $R_{\max}^* \rightarrow 0$.

Portanto, o TQSL é tão bom quanto o teste ótimo (teste da razão da verossimilhança) para detectar a presença de variáveis anômalas na classe dos k_n -intervalos nos casos $k_n / \log n \ll 1/\rho_n$ e $k_n / \log n \gg 1/\rho_n$, desde que $(\log N_n)/k_n \rightarrow 0$.

4.3.4 Teste da Bondade de Ajuste

Considere a classe dos k_n -conjuntos. Nesse caso, sabemos que o TQS tem o mesmo desempenho do teste ótimo se $k_n^2/n \rightarrow \infty$ quando $n \rightarrow \infty$. Agora, se $k_n^2/n \rightarrow 0$ quando $n \rightarrow \infty$, sabemos que o $R_{\max}(f_{TQS}) \rightarrow 1$ para todo teste f_{TQS} . Além disso, sabemos que $R_{\max}(f_{TRVG}) \rightarrow 0$ desde que $\log(n/k_n) \rightarrow \infty$, $\frac{1}{1-\rho_n} \gg \left(\frac{n}{k_n}\right)^2 \log(n/k_n)$. Pelo Corolário 3 temos que $R^* \rightarrow 1$ quando $n \rightarrow \infty$ desde que $\frac{1}{1-\rho_n} \ll \left(\frac{n}{k_n^2}\right)^2$. Observe que $(n/k_n)^2 \log(n/k_n) \gg (n/k_n^2)^2$ quando $k_n^2/n \rightarrow 0$. Desse modo, com base nos testes introduzidos até aqui, não sabemos o que acontece com o risco minimax no regime $k_n^2/n \rightarrow 0$ e $\left(\frac{n}{k_n^2}\right)^2 \ll \frac{1}{1-\rho_n} \ll \left(\frac{n}{k_n}\right)^2 \log(n/k_n)$. Por isso, nesta subseção, introduzimos o teste da bondade de ajuste (TBA) que nos permite deduzir o comportamento do risco minimax no regime acima, complementando os resultados anteriores.

Seja $H_i = \Phi(X_i)$, onde Φ é a função de distribuição da normal padrão. Desse modo, sob \mathbb{P}_0 , temos que as H_i 's são independentes e seguem distribuição uniforme em $(0, 1)$. Sob \mathbb{P}_S , os X_i 's com $i \in S$ estão próximos uns dos outros se nos colocarmos no regime onde $\rho_n \rightarrow 1$ quando $n \rightarrow \infty$. Mais precisamente, temos o seguinte.

Lema 11 (Arias-Castro et al. (2012)). *Suponha que as variáveis aleatórias X_1, \dots, X_{k_n} tenham média zero e variância unitária satisfazendo $\text{Cov}(X_i, X_j) \geq \rho_n > 0$, para todo $i \neq j$. Seja \bar{X} a sua média amostral. Então, para todo $t > 0$,*

$$\mathbb{P} \left\{ \left| \left\{ i : |X_i - \bar{X}| > t \right\} \right| \geq k_n/2 \right\} \leq \frac{2(1 - \rho_n)}{t^2}.$$

Demonstração. Veja que,

$$\begin{aligned} \mathbb{P} \left\{ \left| \left\{ i : |X_i - \bar{X}| > t \right\} \right| \geq k_n/2 \right\} &= \mathbb{P} \left\{ \frac{1}{k_n} \left| \left\{ i : |X_i - \bar{X}| > t \right\} \right| \geq 1/2 \right\} \\ &= \mathbb{P} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{1}_{\{|X_i - \bar{X}| > t\}} \geq 1/2 \right\}. \end{aligned}$$

Se $\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{1}_{\{|X_i - \bar{X}| > t\}} \geq 1/2$, então

$$\begin{aligned} \frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 &\geq \frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \mathbb{1}_{\{|X_i - \bar{X}| > t\}} \\ &\geq \frac{1}{k_n} \sum_{i=1}^{k_n} t^2 \mathbb{1}_{\{|X_i - \bar{X}| > t\}} \\ &\geq t^2/2. \end{aligned}$$

Logo,

$$\mathbb{P} \left\{ \left| \left\{ i : |X_i - \bar{X}| > t \right\} \right| \geq k_n/2 \right\} \leq \mathbb{P} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \geq t^2/2 \right\}.$$

Veja que,

$$\frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 = \frac{1}{k_n} \sum_{i=1}^{k_n} X_i^2 - \bar{X}^2.$$

Desse modo,

$$\begin{aligned} \mathbb{E} \left[\frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \right] &= \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E} [X_i^2] - \mathbb{E} [\bar{X}^2] \\ &= 1 - \frac{1}{k_n^2} \mathbb{E} \left[\sum_{i=1}^{k_n} X_i^2 + \sum_{i \neq j} X_i X_j \right] \\ &= 1 - \frac{1}{k_n} - \frac{1}{k_n^2} \sum_{i \neq j} \mathbb{E} [X_i X_j] \\ &\stackrel{(i)}{\leq} 1 - \frac{1}{k_n} - \frac{1}{k_n^2} k_n(k_n - 1) \rho_n \\ &= \left(1 - \frac{1}{k_n} \right) (1 - \rho_n) \\ &\leq (1 - \rho_n), \end{aligned}$$

ond em (i) usamos as hipótese do Lema 11. Logo, pela desigualdade de Markov temos que

$$\mathbb{P} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \geq t^2/2 \right\} \leq \frac{2}{t^2} \mathbb{E} \left[\frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \right] \leq \frac{2(1 - \rho_n)}{t^2}.$$

Portanto

$$\begin{aligned} \mathbb{P} \left\{ \left| \left\{ i : |X_i - \bar{X}| > t \right\} \right| \geq k_n/2 \right\} &\leq \mathbb{P} \left\{ \frac{1}{k_n} \sum_{i=1}^{k_n} (X_i - \bar{X})^2 \geq t^2/2 \right\} \\ &\leq \frac{2(1 - \rho_n)}{t^2}, \end{aligned}$$

concluindo a prova. \square

A ideia é detectar grandes concentrações não usuais de H_i 's, que é uma forma de teste de bondade de ajuste para a distribuição uniforme. A partir do Lema 11 vemos que a concentração (em torno da média amostral) ocorre em um intervalo de tamanho um pouco menor que $\sqrt{1 - \rho_n}$, pois nesse caso, para $\epsilon > 0$ fixado temos que

$$\mathbb{P} \left\{ \left| \left\{ i : |X_i - \bar{X}| > (1 - \rho_n)^{\frac{1}{2} - \epsilon} \right\} \right| \geq k_n/2 \right\} \leq 2(1 - \rho_n)^{2\epsilon},$$

que converge à zero quando $\rho_n \rightarrow 1$.

Escolha um inteiro m tal que $m \gg (n/k_n^2) \log n$ e particione o intervalo $[0, 1)$ em m intervalos de tamanho $1/m$, denotados por I_s com $s = 1, \dots, m$. Mais precisamente, consideramos $I_s = [\frac{s-1}{m}, \frac{s}{m})$ com $s = 1, \dots, m$. Seja $B_s = |\{i : H_i \in I_s\}|$ para cada $s = 1, \dots, m$. O teste de bondade de ajuste é definido por

$$f_{TBA}(X) = 0 \text{ se e somente se } \max_{s=1, \dots, m} B_s < t_n,$$

onde t_n é um regime apropriado.

Proposição 18 (Arias-Castro et al. (2012)). *Seja \mathcal{C}_n a classe de todos os k_n -conjuntos no caso em que $k_n^2/n \rightarrow 0$ e $k_n/\log n \rightarrow \infty$. No TBA acima, escolha m tal que $(n/k_n^2) \log n \ll m \ll n/\log n$. Quando $(1 - \rho_n)^{1/2} \ll 1/m$, o risco do pior caso possível para o TBA com $t_n = n/m + \sqrt{(3n \log m)/m}$ converge à zero.*

Demonstração. Veja que,

$$R_{\max}(f_{TBA}) = \mathbb{P}_0 \left\{ \max_{s=1, \dots, m} B_s \geq t_n \right\} + \max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s < t_n \right\}.$$

Primeiramente, vamos dar cota superior para $\mathbb{P}_0 \left\{ \max_{s=1, \dots, m} B_s \geq t_n \right\}$.

Observe que pela cota da união temos que

$$\begin{aligned}\mathbb{P}_0 \left\{ \max_{s=1,\dots,m} B_s \geq t_n \right\} &= \mathbb{P}_0 \left\{ \cup_{s=1,\dots,m} B_s \geq t_n \right\} \\ &\leq \sum_{s=1}^m \mathbb{P}_0 \{ B_s \geq t_n \} \\ &= m \mathbb{P}_0 \{ B_1 \geq t_n \},\end{aligned}$$

pois $B_s \sim \text{Bin}(n, 1/m)$ para todo $s = 1, \dots, m$. Como $B_1 = \sum_{i=1}^n Y_i$ onde $Y_i \sim \text{Bern}(1/m)$ e $|Y_i - 1/m| \leq 1$, logo pela desigualdade de Bernstein (Apêndice D) temos que para todo $t_n > 0$,

$$\mathbb{P}_0 \left\{ B_1 - \frac{n}{m} \geq t_n \right\} \leq \exp \left(\frac{-t_n^2}{2 \left(\frac{n}{m} \left(1 - \frac{1}{m} \right) + \frac{t_n}{3} \right)} \right).$$

Tomando $t_n = \sqrt{(3n \log m)/m}$ temos que

$$\mathbb{P}_0 \left\{ B_1 \geq \frac{n}{m} + \sqrt{\frac{3n \log m}{m}} \right\} \leq \exp \left(\frac{-3n \log m}{2 \left(n \left(1 - \frac{1}{m} \right) + \sqrt{\frac{nm \log m}{3}} \right)} \right).$$

Logo, pela cota da união temos que

$$\begin{aligned}\mathbb{P}_0 \left\{ \max_{s=1,\dots,m} B_s \geq t_n \right\} &\leq m \mathbb{P}_0 \{ B_1 \geq t_n \} \\ &\leq \exp \left(\frac{-3n \log m}{2 \left(n \left(1 - \frac{1}{m} \right) + \sqrt{\frac{nm \log m}{3}} \right)} + \log m \right),\end{aligned}$$

que converge à zero se

$$\frac{-3n \log m}{2 \left(n \left(1 - \frac{1}{m} \right) + \sqrt{\frac{nm \log m}{3}} \right)} + \log m \rightarrow -\infty.$$

Veja que

$$\frac{-3n \log m}{2 \left(n \left(1 - \frac{1}{m} \right) + \sqrt{\frac{nm \log m}{3}} \right)} + \log m = -\log m \left(\frac{3}{2 \left(\left(1 - \frac{1}{m} \right) + \sqrt{\frac{m \log m}{3n}} \right)} - 1 \right),$$

que tende à $-\infty$ quando $n \rightarrow \infty$ pois como $(m \log n)/n \rightarrow 0$, logo $\frac{m \log m}{n} = \frac{m \log n}{n} \times \frac{\log m}{\log n} \rightarrow 0$. Portanto, $\mathbb{P}_0 \{ \max_{s=1, \dots, m} B_s \geq t_n \} \rightarrow 0$ quando $n \rightarrow \infty$.

Agora, vamos dar cota inferior para o complementar da probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s < t_n \right\}.$$

Defina

$$I = \left\{ i \in S : |X_i - \bar{X}_S| \leq 1/m \right\} \text{ e } \bar{X}_S = \frac{1}{k_n} \sum_{i \in S} X_i.$$

Embora o conjunto I seja aleatório, pelo Lema 11 temos que

$$\begin{aligned} \mathbb{P}_S \{ |I| \geq k_n/2 \} &= \mathbb{P}_S \{ |I^c| < k_n/2 \} \\ &\geq 1 - \mathbb{P}_S \{ |I^c| \geq k_n/2 \} \\ &\geq 1 - 2(1 - \rho_n)m^2. \end{aligned}$$

Como $(1 - \rho_n)^{1/2}m \rightarrow 0$ por hipótese, logo $\mathbb{P}_S \{ |I| \geq k_n/2 \} \rightarrow 1$.

Suponha que $\bar{X}_{s_0} \in I_{s_0}$ para algum $s_0 \in \{1, \dots, m\}$. No evento $|I| \geq k_n/2$,

$$\begin{aligned} |H_i - \Phi(\bar{X}_{s_0})| &= |\Phi(X_i) - \Phi(\bar{X}_{s_0})| \\ &= \Phi'(\psi_i) |X_i - \bar{X}_{s_0}| \\ &\leq \frac{1}{\sqrt{2\pi}} |X_i - \bar{X}_{s_0}|, \end{aligned}$$

onde ψ_i está entre X_i e \bar{X}_{s_0} . Se $i \in I$, então

$$|H_i - \Phi(\bar{X}_{s_0})| \leq \frac{1}{m\sqrt{2\pi}} \leq \frac{1}{m}.$$

Se $\Phi(\bar{X}_{s_0}) \in I_{s_0}$ e $i \in I$, então $H_i \in I_{s_0-1} \cup I_{s_0} \cup I_{s_0+1}$, em outras palavras, este $i \in I$ vai ser contado em I_{s_0-1} ou I_{s_0} ou I_{s_0+1} . Logo $I \subseteq I_{s_0-1} \cup I_{s_0} \cup I_{s_0+1}$.

Ainda no evento $|I| \geq k_n/2$, temos que $|B_{s_0-1}| + |B_{s_0}| + |B_{s_0+1}| \geq |I| \geq k_n/2$. Assim, $|B_l| \geq k_n/6$ para algum $l \in \{s_0 - 1, s_0, s_0 + 1\}$. Veja que,

$$\mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s \geq t \right\} \geq \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s \geq t, |I| \geq \frac{k_n}{2} \right\}.$$

Seja $\tilde{B}_s = |\{i \in \{1, \dots, n\} \setminus S : H_i \in I_s\}|$, isto é, a contagem dos não anômalos H_i que estão em I_s . Note que, sob \mathbb{P}_S , $\tilde{B}_s \sim \text{Bin}(n - k_n, 1/m)$.

Seja $\tilde{t}_n = (n - k_n)/m - \sqrt{(3n \log m)/m}$. Logo, pela desigualdade de Bernstein (Apêndice D) e a cota da união temos que

$$\mathbb{P}_S \left\{ \max_{s=1, \dots, m} \tilde{B}_s \leq \tilde{t}_n \right\} \leq \exp \left(\log m - \frac{3 \log m}{2 \left(\frac{1 - k_n/n}{m} \left(1 - \frac{1}{m} \right) + \sqrt{\frac{m \log(m)}{3n}} \right)} \right),$$

que converge à zero quando $n \rightarrow \infty$ pois $m \log(n)/n \rightarrow 0$ e $k_n^2/n \rightarrow 0$.

Veja que $B_s = \tilde{B}_s + |\{i \in S : H_i \in I_s\}|$. Logo, no evento $|I| \geq k_n/2$, existe $s \in \{1, \dots, m\}$ tal que $|\{i \in S : H_i \in I_s\}| \geq k_n/6$. Então,

$$\begin{aligned} \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s \geq \tilde{t}_n + k_n/6, |I| \geq k_n/2 \right\} &\geq \mathbb{P}_S \left\{ \max_{s=1, \dots, m} \tilde{B}_s \geq \tilde{t}_n \right\} - \mathbb{P}_S \{ |I| \leq k_n/2 \} \\ &\geq \mathbb{P}_S \left\{ \max_{s=1, \dots, m} \tilde{B}_s \geq \tilde{t}_n \right\} \\ &\quad + \mathbb{P}_S \{ |I| \geq k_n/2 \} - 1 \\ &\geq 1 - \mathbb{P}_S \left\{ \max_{s=1, \dots, m} \tilde{B}_s \leq \tilde{t}_n \right\} - 2(1 - \rho_n)m^2, \end{aligned}$$

que converge à 1 quando $n \rightarrow \infty$. Agora,

$$\begin{aligned} \tilde{t}_n + \frac{k_n}{6} - t_n &= \frac{k_n}{6} - \frac{k_n}{m} - 2\sqrt{\frac{3n \log m}{m}} \\ &= k_n \left(\frac{1}{6} - \frac{1}{m} - 2\sqrt{\frac{3n \log m}{k_n^2 m}} \right). \end{aligned}$$

Como $(n \log n)/k_n^2 m \rightarrow 0$ por hipótese, logo $(n \log m)/k_n^2 m \rightarrow 0$ quando $n \rightarrow \infty$ e temos que $\tilde{t}_n + \frac{k_n}{6} - t_n \geq 0$. Portanto,

$$\mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s \geq t_n + k_n/6, |I| \geq k_n/2 \right\} \geq \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s \geq \tilde{t}_n + k_n/6, |I| \geq k_n/2 \right\},$$

que converge à 1 quando $n \rightarrow \infty$.

Como vale para todo $S \in \mathcal{C}_n$, logo $\max_{S \in \mathcal{C}_n} \mathbb{P}_S \{ \max_{s=1, \dots, m} B_s \geq t_n \} \rightarrow 1$, implicando que a probabilidade

$$\max_{S \in \mathcal{C}_n} \mathbb{P}_S \left\{ \max_{s=1, \dots, m} B_s < t_n \right\} \rightarrow 0,$$

quando $n \rightarrow \infty$. □

No caso dos k_n -conjuntos e no regime $k_n^2/n \rightarrow 0$ e $k_n/\log n \rightarrow \infty$ temos que:

- Se $1/(1-\rho_n) \ll n^2/k_n^4$, então pelo Corolário 3 temos que $R_{\max}^* \geq R^* \rightarrow 1$. Segue que $R_{\max}(f_{TBA}) \rightarrow 1$.
- Veja que $n^2(\log n)^2/k_n^4 \ll 1/(1-\rho_n)$, é equivalente à, $(n \log n)/k_n^2 \ll 1/(1-\rho_n)^{1/2}$. Desse modo, se $(n \log n)/k_n^2 \ll 1/(1-\rho_n)^{1/2}$ vale, então posso escolher m satisfazendo as hipóteses da Proposição 18 e concluimos que $R_{\max}^* \leq R_{\max}(f_{TBA}) \rightarrow 0$.

Para verificar que podemos escolher m no segundo item do modo que satisfaça as hipóteses da Proposição 18, note primeiramente que m precisa satisfazer

$$\frac{n \log n}{k_n^2} \ll m \ll \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\}. \quad (4.11)$$

É fácil ver que $(n \log n)/k_n^2 \ll n/\log n$ se $k_n/\log n \rightarrow \infty$ e se além disso, $(n \log n)/k_n^2 \ll 1/(1-\rho_n)^{1/2}$, então

$$\frac{n \log n}{k_n^2} \ll \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\}, \quad (4.12)$$

e portanto faz sentido determinar m satisfazendo (4.11).

Como procuramos m tal que $m \ll \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\}$, este m deve ser escrito como $m = \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\} a_n$, onde $a_n \rightarrow 0$. Por outro lado, por (4.12), existe b_n tal que $(n \log n)/k_n^2 = \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\} b_n$, com $b_n \rightarrow 0$. Assim observando que,

$$\begin{aligned} \frac{n \log n}{k_n^2} &= \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\} \times b_n \\ &= \min \left\{ \frac{1}{(1-\rho_n)^{1/2}}, \frac{n}{\log n} \right\} a_n \times \frac{b_n}{a_n} \\ &= m \times \frac{b_n}{a_n}, \end{aligned}$$

concluimos que se $b_n/a_n \rightarrow 0$, então m satisfaz (4.11).

Portanto, esquecendo o fator logaritmo no segundo item, temos que o TBA é tão bom quanto o teste ótimo (teste da razão de verossimilhança) para detectar a presença de variáveis anômalas na classe dos k_n -intervalos nos casos $n^2/k_n^4 \ll 1/(1-\rho_n)$ e $n^2/k_n^4 \gg 1/(1-\rho_n)$, desde que $k_n^2/n \rightarrow 0$ e $k_n/\log n \rightarrow \infty$.

Apêndice A

Lema de Neyman-Pearson

Lema 12 (Neyman-Pearson). *Sejam \mathbb{Q}_0 e \mathbb{Q}_1 duas medidas de probabilidade em \mathbb{R}^n com densidades q_0 e q_1 em relação a uma medida σ -finita ν em \mathbb{R}^n . Defina $f_{TRV} : \mathbb{R}^n \rightarrow \{0, 1\}$ para todo $y \in \mathbb{R}^n$ por*

$$f_{TRV}(y) = \begin{cases} 1, & \text{se } q_0(y) \leq q_1(y), \\ 0, & \text{caso contrário.} \end{cases}$$

Então, para toda função $f : \mathbb{R}^n \rightarrow \{0, 1\}$ mensurável vale que

$$\mathbb{Q}_0\{y \in \mathbb{R}^n : f(y) = 1\} + \mathbb{Q}_1\{y \in \mathbb{R}^n : f(y) = 0\} \geq \int_{\mathbb{R}^n} \min(q_0(y), q_1(y)) d\nu(y).$$

A igualdade é satisfeita para o teste f_{TRV} .

Demonstração. Para encurtar a notação, para qualquer função $g : \mathbb{R}^n \rightarrow \mathbb{R}$ integrável com relação a medida ν e $A \subset \mathbb{R}^n$ mensurável, vamos denotar

$$I(A, g) = \int_A g(y) d\nu(y),$$

e além disso, $\{y \in \mathbb{R}^n : g(y) > c\} = \{g > c\}$, $\{y \in \mathbb{R}^n : g(y) \leq c\} = \{g \leq c\}$ e $\{y \in \mathbb{R}^n : g(y) = c\} = \{g = c\}$ para todo $c \in \mathbb{R}$. Para qualquer teste $f : \mathbb{R}^n \rightarrow \{0, 1\}$, defina a região de rejeição $R_f = \{f = 1\}$.

Por um lado, veja que

$$\begin{aligned} \mathbb{Q}_0\{f_{TRV} = 1\} + \mathbb{Q}_1\{f_{TRV} = 0\} &= I(\{f_{TRV} = 1\}, q_0) + I(\{f_{TRV} = 0\}, q_1) \\ &= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_0\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_1\right). \end{aligned}$$

Note que a expressão $I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_0\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_1\right)$ pode ser reescrita como

$$I\left(\left\{\frac{q_1}{q_0} > 1\right\}, \min(q_0, q_1)\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, \min(q_0, q_1)\right) = I(\mathbb{R}^n, \min(q_0, q_1)).$$

Portanto, concluímos que $\mathbb{Q}_0\{f_{TRV} = 1\} + \mathbb{Q}_1\{f_{TRV} = 0\} = I(\mathbb{R}^n, \min(q_0, q_1))$. Dessa maneira, para concluir o resultado precisamos mostrar que

$$\mathbb{Q}_0\{f = 1\} + \mathbb{Q}_1\{f = 0\} \geq I(\mathbb{R}^n, \min(q_0, q_1)).$$

Para esse fim, denote $R^* = \left\{\frac{q_1}{q_0} > 1\right\}$, a região de rejeição do teste de razão de verossimilhança f_{TRV} e observe que para qualquer teste f

$$\begin{aligned} \mathbb{Q}_0\{f = 1\} + \mathbb{Q}_1\{f = 0\} &= \mathbb{Q}_0\{f = 1\} + 1 - \mathbb{Q}_1\{f = 1\} \\ &= \mathbb{Q}_0\{R_f\} + 1 - \mathbb{Q}_1\{R_f\} \\ &= 1 + I(R_f, q_0) - I(R_f, q_1) \\ &= 1 + I(R_f, q_0 - q_1). \end{aligned}$$

Agora, observe que $1 + I(R_f, q_0 - q_1)$ é igual a

$$\begin{aligned} &1 + I(R_f \cap R^*, q_0 - q_1) + I(R_f \cap (R^*)^c, q_0 - q_1) \\ &= 1 - I(R_f \cap R^*, |q_0 - q_1|) + I(R_f \cap (R^*)^c, |q_0 - q_1|) \\ &= 1 - I(\mathbb{R}^n, |q_0 - q_1| \mathbb{1}_{\{R_f \cap R^*\}}) + I(\mathbb{R}^n, |q_0 - q_1| \mathbb{1}_{\{R_f \cap (R^*)^c\}}) \\ &= 1 + I\left(\mathbb{R}^n, |q_0 - q_1| \left(\mathbb{1}_{\{R_f \cap R^*\}} - \mathbb{1}_{\{R_f \cap (R^*)^c\}}\right)\right). \end{aligned}$$

Além disso, veja que para todo R_f vale que

$$\mathbb{1}_{\{R_f \cap (R^*)^c\}} \geq 0 \quad \text{e} \quad \mathbb{1}_{\{R_f \cap R^*\}} \leq \mathbb{1}_{\{R^*\}},$$

de modo que

$$\mathbb{1}_{\{R_f \cap (R^*)^c\}} - \mathbb{1}_{\{R_f \cap R^*\}} \geq -\mathbb{1}_{\{R^*\}} \quad \text{para todo } R_f.$$

Portanto, concluímos que para todo teste f

$$\begin{aligned} \mathbb{Q}_0\{f = 1\} + \mathbb{Q}_1\{f = 0\} &\geq 1 - I(\mathbb{R}^n, |q_0 - q_1| \mathbb{1}_{\{R^*\}}) \\ &= 1 - I(R^*, |q_0 - q_1|). \end{aligned}$$

Observando que

$$I(\mathbb{R}^n, |q_0 - q_1|) = I\left(\left\{\frac{q_0}{q_1} > 1\right\}, q_0 - q_1\right) + I\left(\left\{\frac{q_0}{q_1} \leq 1\right\}, q_1 - q_0\right),$$

e como

$$\begin{aligned}
I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_0 - q_1\right) &= I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_0\right) - I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_1\right) \\
&= 1 - I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_0\right) - \left(1 - I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_1\right)\right) \\
&= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_1 - q_0\right) \\
&= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, |q_1 - q_0|\right),
\end{aligned}$$

temos que

$$I(\mathbb{R}^n, |q_0 - q_1|) = 2I\left(\left\{\frac{q_1}{q_0} > 1\right\}, |q_1 - q_0|\right) = I(R^*, |q_1 - q_0|),$$

e assim,

$$\mathbb{Q}_0\{f = 1\} + \mathbb{Q}_1\{f = 0\} \geq 1 - \frac{I(\mathbb{R}^n, |q_0 - q_1|)}{2}.$$

Para concluir a demonstração, resta mostrar que

$$I(\mathbb{R}^n, |q_0 - q_1|) = 2 - 2I(\mathbb{R}^n, \min(q_0, q_1)).$$

Para esse fim note que,

$$\begin{aligned}
I(\mathbb{R}^n, |q_0 - q_1|) &= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_1 - q_0\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_0 - q_1\right) \\
&= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_1\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_0\right) \\
&\quad - \left(I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_0\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_1\right)\right) \\
&= I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_1\right) + I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_0\right) - I(\mathbb{R}^n, \min(q_0, q_1)) \\
&= 2 - I\left(\left\{\frac{q_1}{q_0} \leq 1\right\}, q_1\right) - I\left(\left\{\frac{q_1}{q_0} > 1\right\}, q_0\right) - I(\mathbb{R}^n, \min(q_0, q_1)) \\
&= 2 - 2I(\mathbb{R}^n, \min(q_0, q_1)),
\end{aligned}$$

concluindo a demonstração. \square

Apêndice B

Lema de Slepian

Lema 13. (*Desigualdade de Slepian, forma funcional*) Considere dois vetores aleatórios Gaussianos X e Y em \mathbb{R}^n com média zero. Assuma que para todo $i, j = 1, \dots, n$, nós temos que

$$\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2] \quad e \quad \mathbb{E}[X_i X_j] \leq \mathbb{E}[Y_i Y_j].$$

Considere a função duas vezes diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0 \quad \text{para todo } i \neq j.$$

Então,

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)].$$

Para demonstrar o lema acima, vamos precisar dos seguintes resultados.

Lema 14. (*Integração Gaussiana multivariada por partes*) Seja $X \sim \mathcal{N}(0, \Sigma)$. Então para qualquer função diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que as esperanças abaixo estejam bem definidas, temos que

$$\mathbb{E}[X f(X)] = \Sigma \mathbb{E}[\nabla f(X)].$$

Demonstração. Vamos provar somente o caso em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tem suporte compacto, isto é, $f(x) = 0$ para $\|x\|_2 > M$ para algum $M > 0$. O resultado da integração vale em geral e segue por um argumento de aproximação.

Veja que a igualdade que queremos mostrar é equivalente à

$$\mathbb{E}[X_i f(X)] = \sum_{j=1}^n \Sigma_{ij} \mathbb{E}\left[\frac{\partial f}{\partial x_j}(X)\right],$$

para $i = 1, \dots, n$. A função de densidade de X é dada por

$$\phi(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} \sum_{i,j=1}^n x_i x_j (\Sigma^{-1})_{ij} \right).$$

Veja que para todo $k = 1, \dots, n$ temos que

$$\begin{aligned} \frac{\partial}{\partial x_k} \phi(x) &= \phi(x) \left(-\sum_{j=1}^n x_j (\Sigma^{-1})_{kj} \right) \\ &= \phi(x) (-\Sigma^{-1}x)_k. \end{aligned}$$

Logo, $\nabla \phi(x) = \phi(x) (-\Sigma^{-1}x) = -\Sigma^{-1} \phi(x)x$. Assim, $\Sigma \nabla \phi(x) = -\phi(x)x$ é equivalente à

$$\sum_{j=1}^n \Sigma_{ij} \frac{\partial \phi(x)}{\partial x_j} = -\phi(x)x_i \quad \text{para } i = 1, \dots, n.$$

Observe que,

$$\begin{aligned} \mathbb{E} [X_i f(X)] &= \int_{\mathbb{R}^n} x_i f(x) \phi(x) dx \\ &= \int_{\mathbb{R}^n} f(x) \phi(x) x_i dx \\ &= \sum_{j=1}^n \Sigma_{ij} \left(-\int_{\mathbb{R}^n} f(x) \frac{\partial \phi(x)}{\partial x_j} dx \right). \end{aligned}$$

Por Fubini,

$$\int_{\mathbb{R}^n} f(x) \frac{\partial \phi(x)}{\partial x_j} dx = \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} f(x) \frac{\partial \phi(x)}{\partial x_j} dx_j \right) d\tilde{x},$$

onde $d\tilde{x} = dx_1 \dots dx_n$ para todo $n \neq j$. Como a função f tem suporte compacto, então aplicando a fórmula de integração por partes temos que

$$\int_{\mathbb{R}} f(x) \frac{\partial \phi(x)}{\partial x_j} dx_j = - \int_{\mathbb{R}} \frac{\partial f(x)}{\partial x_j} \phi(x) dx_j.$$

Assim,

$$\begin{aligned} - \int_{\mathbb{R}^n} f(x) \frac{\partial \phi(x)}{\partial x_j} dx &= \int_{\mathbb{R}^n} \frac{\partial f(x)}{\partial x_j} \phi(x) dx_j \\ &= \mathbb{E} \left[\frac{\partial f}{\partial x_j}(X) \right], \end{aligned}$$

e concluimos que para $i = 1, \dots, n$,

$$\mathbb{E} [X_i f(X)] = \sum_{j=1}^n \Sigma_{ij} \mathbb{E} \left[\frac{\partial f}{\partial x_j}(X) \right].$$

□

Lema 15. (*Interpolação Gaussiana*) Considere dois vetores Gaussianos independentes $X \sim \mathcal{N}(0, \Sigma^X)$ e $Y \sim \mathcal{N}(0, \Sigma^Y)$. Defina o vetor Gaussiano interpolado

$$Z(u) = \sqrt{u}X + \sqrt{1-u}Y, \quad u \in (0, 1).$$

Então para qualquer função duas vezes diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que as esperanças abaixo estejam bem definidas, temos que

$$\frac{d}{du} \mathbb{E} [f(Z(u))] = \frac{1}{2} \sum_{i,j=1}^n (\Sigma_{ij}^X - \Sigma_{ij}^Y) \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)) \right].$$

Demonstração. Pela regra da cadeia multivariada nós temos que

$$\begin{aligned} \frac{d}{du} \mathbb{E} [f(Z(u))] &= \sum_{i=1}^n \mathbb{E} \left[\frac{\partial f}{\partial x_i}(Z(u)) \frac{dZ_i}{du} \right] \\ &= \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial f}{\partial x_i}(Z(u)) \left(\frac{X_i}{\sqrt{u}} - \frac{Y_i}{\sqrt{1-u}} \right) \right]. \end{aligned} \tag{B.1}$$

Vamos separar a soma acima em duas e primeiramente, vamos calcular a contribuição dos termos contendo X_i . Veja que,

$$\mathbb{E} \left[\frac{\partial f}{\partial x_i}(Z(u)) X_i \right] = \mathbb{E}_Y \left[\mathbb{E}_X \left[\frac{\partial f}{\partial x_i}(Z(u)) X_i | Y \right] \right].$$

Seja

$$\psi(Y) = \mathbb{E}_X \left[\frac{\partial f}{\partial x_i}(\sqrt{u}X + \sqrt{1-u}Y) X_i | Y \right].$$

Pelo princípio da substituição,

$$\psi(y) = \mathbb{E}_X \left[\frac{\partial f}{\partial x_i}(\sqrt{u}X + \sqrt{1-u}y) X_i \right],$$

é função de X . Desse modo, pelo Lema 14 temos que

$$\begin{aligned}\psi(y) &= \mathbb{E}_X \left[\frac{\partial f}{\partial x_i} (\sqrt{u}X + \sqrt{1-uy})X_i \right] \\ &= \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E}_X \left[\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} (\sqrt{u}X + \sqrt{1-uy}) \right) \right] \\ &= \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E}_X \left[\frac{\partial^2 f}{\partial x_i \partial x_j} (\sqrt{u}X + \sqrt{1-uy}) \right] \sqrt{u}.\end{aligned}$$

Assim,

$$\begin{aligned}\mathbb{E} \left[\frac{\partial f}{\partial x_i} (Z(u))X_i \right] &= \mathbb{E}_Y [\psi(Y)] \\ &= \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} (Z(u)) \right] \sqrt{u}.\end{aligned}$$

Logo,

$$\frac{1}{\sqrt{u}} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial f}{\partial x_i} (Z(u))X_i \right] = \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} (Z(u)) \right]. \quad (\text{B.2})$$

De maneira análoga, podemos calcular a contribuição dos termos contendo Y_i e segue que

$$\frac{1}{\sqrt{1-u}} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial f}{\partial x_i} (Z(u))Y_i \right] = \sum_{i=1}^n \sum_{j=1}^n \Sigma_{ij}^Y \mathbb{E} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} (Z(u)) \right]. \quad (\text{B.3})$$

Concluimos a demonstração substituindo (B.2) e (B.3) em (B.1). \square

Demonstração do Lema 13. As hipóteses implicam que as entradas das matrizes de covariância Σ^X e Σ^Y de X e Y satisfazem

$$\Sigma_{ii}^X = \Sigma_{ii}^Y \quad \text{e} \quad \Sigma_{ij}^X \leq \Sigma_{ij}^Y,$$

para todo $i, j = 1, \dots, n$. Podemos assumir que X e Y são independentes. Aplicando o Lema 15 e as hipóteses, concluimos que

$$\frac{d}{du} \mathbb{E} [f(Z(u))] \geq 0.$$

Logo, $\mathbb{E} [f(Z(u))]$ é crescente em u . Veja que

$$\mathbb{E} [f(Z(1))] = \mathbb{E} [f(X)] \quad \text{e} \quad \mathbb{E} [f(Z(0))] = \mathbb{E} [f(Y)].$$

Portanto

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)],$$

concluindo a prova.

□

Apêndice C

Lema de Tsirelson

Lema 16. (Tsirelson) Seja $Y = (Y_1, \dots, Y_n)$ vetor de n variáveis aleatórias independentes com distribuição normal padrão. Seja $g : \mathbb{R}^n \rightarrow \mathbb{R}$ função Lipschitz com constante de Lipschitz K (com respeito a distância Euclidiana). Então para todo $t > 0$,

$$\mathbb{P}\{g(Y) - \mathbb{E}[g(Y)] \geq t\} \leq e^{-t^2/2K}.$$

Para demonstrar o lema acima, vamos precisar do seguinte resultado.

Lema 17. Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável. Então para qualquer função convexa $\phi : \mathbb{R} \rightarrow \mathbb{R}$ nós temos que

$$\mathbb{E} \left[\phi \left(f(X) - \mathbb{E} [f(X)] \right) \right] \leq \mathbb{E} \left[\phi \left(\frac{\pi}{2} \langle \nabla f(X), Y \rangle \right) \right],$$

onde $X, Y \sim \mathcal{N}(0, I_n)$ são normais padrão multivariadas e independentes.

Demonstração. Para cada $\theta \in [0, \pi/2]$, considere o vetor aleatória $Z(\theta) \in \mathbb{R}^n$ com componentes,

$$Z_k(\theta) = X_k \sin(\theta) + Y_k \cos(\theta),$$

para $k = 1, 2, \dots, n$. Como ϕ é convexa, pela desigualdade de Jensen temos que

$$\begin{aligned} \mathbb{E}_X \left[\phi \left(f(X) - \mathbb{E}_Y [f(Y)] \right) \right] &= \mathbb{E}_X \left[\phi \left(\mathbb{E}_Y [f(X) - f(Y)] \right) \right] \\ &\leq \mathbb{E}_X \left[\mathbb{E}_Y \left[\phi \left(f(X) - f(Y) \right) \right] \right] \quad (\text{C.1}) \\ &= \mathbb{E}_{X,Y} \left[\phi \left(f(X) - f(Y) \right) \right], \end{aligned}$$

onde \mathbb{E}_X denota o valor esperado com respeito a X . Como $Z_k(0) = Y_k$ e $Z_k(\pi/2) = X_k$ para todo $k = 1, \dots, n$, nós temos que

$$f(X) - f(Y) = \int_0^{\pi/2} \frac{d}{d\theta} f(Z(\theta)) d\theta = \int_0^{\pi/2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle d\theta, \quad (\text{C.2})$$

onde $Z'(\theta) \in \mathbb{R}^n$ denota o vetor de derivadas cujas componentes são

$$Z'_k(\theta) = X_k \cos(\theta) - Y_k \sin(\theta),$$

para todo $k = 1, \dots, n$. Substituindo (C.2) na última igualdade de (C.1) temos que

$$\begin{aligned} \mathbb{E}_X \left[\phi \left(f(X) - \mathbb{E}_Y [f(Y)] \right) \right] &\leq \mathbb{E}_{X,Y} \left[\phi \left(\int_0^{\pi/2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle d\theta \right) \right] \\ &= \mathbb{E}_{X,Y} \left[\phi \left(\frac{1}{\pi/2} \int_0^{\pi/2} \frac{\pi}{2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle d\theta \right) \right] \\ &\leq \frac{1}{\pi/2} \int_0^{\pi/2} \mathbb{E}_{X,Y} \left[\phi \left(\frac{\pi}{2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle \right) \right] d\theta, \end{aligned}$$

pois ϕ é função convexa. Veja que para cada $\theta \in [0, \pi/2]$,

$$\text{Cov}(Z_k(\theta), Z'_k(\theta)) = 0, \quad \text{Var}(Z_k(\theta)) = 1, \quad \text{Var}(Z'_k(\theta)) = 1,$$

e como $Z_k(\theta)$ e $Z'_k(\theta)$ seguem distribuição normal, temos que $(Z_k(\theta), Z'_k(\theta)) \sim \mathcal{N}(0, I_2)$ e não depende de θ . Portanto,

$$\frac{1}{\pi/2} \int_0^{\pi/2} \mathbb{E}_{X,Y} \left[\phi \left(\frac{\pi}{2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle \right) \right] d\theta = \mathbb{E} \left[\phi \left(\frac{\pi}{2} \langle \nabla f(\tilde{X}), \tilde{Y} \rangle \right) \right],$$

onde (\tilde{X}, \tilde{Y}) são vetores aleatórios n dimensionais com distribuição normal padrão. □

Demonstração do Lema 16. Veja que

$$\begin{aligned} \mathbb{P}\{g(Y) - \mathbb{E}[g(Y)] \geq t\} &= \mathbb{P}\{\lambda(g(Y) - \mathbb{E}[g(Y)]) \geq \lambda t\} \\ &= \mathbb{P}\{\exp(\lambda(g(Y) - \mathbb{E}[g(Y)])) \geq \exp(\lambda t)\} \\ &\leq \mathbb{E} \left[\exp(\lambda(g(Y) - \mathbb{E}[g(Y)])) \right] / e^{-\lambda t}, \end{aligned}$$

pela desigualdade de Markov.

Veja que para $\lambda \in \mathbb{R}$ fixo, $e^{\lambda t}$ é função convexa em t . Denote por \mathbb{E}_Y , o valor esperado com respeito a variável aleatória Y e $\mathbb{E}_{X,Y}$, o valor esperado com respeito as variáveis aleatórias X e Y . Logo pelo Lema 17 temos que

$$\begin{aligned} \mathbb{E}_Y \left[\exp(\lambda(g(Y) - \mathbb{E}[g(Y)])) \right] &\leq \mathbb{E}_{X,Y} \left[\exp\left(\frac{\pi}{2} \langle \nabla g(Y), X \rangle\right) \right] \\ &= \mathbb{E}_Y \left[\mathbb{E}_X \left[\exp\left(\frac{\pi}{2} \langle \nabla g(Y), X \rangle\right) \right] \right]. \end{aligned}$$

Como, dado Y , $\langle \nabla g(Y), X \rangle \sim \mathcal{N}(0, \|\nabla g(Y)\|_2^2)$ e usando a função geradora de momentos da distribuição normal, concluímos que

$$\mathbb{E}_Y \left[\exp(\lambda(g(Y) - \mathbb{E}[g(Y)])) \right] \leq \mathbb{E}_Y \left[\exp\left(\frac{\lambda^2 \pi^2}{8} \|\nabla g(Y)\|_2^2\right) \right].$$

Como g é função Lipschitz, logo $\|\nabla g(y)\|_2 \leq \sqrt{K}$ para todo $y \in \mathbb{R}^n$ e então

$$\mathbb{E} \left[\exp\left(\lambda(g(Y) - \mathbb{E}[g(Y)])\right) \right] \leq e^{\frac{1}{8} \lambda^2 \pi^2 K}.$$

Assim,

$$\begin{aligned} \mathbb{P}\{g(Y) - \mathbb{E}[g(Y)] \geq t\} &\leq \inf_{\lambda > 0} \exp\left\{\lambda t + \frac{1}{8} \lambda^2 \pi^2 K\right\} \\ &= e^{-\frac{2t^2}{\pi^2 K}}, \end{aligned}$$

pois $\lambda = \frac{4t}{\lambda^2 \pi^2 K}$ é ponto de mínimo de $g(\lambda) = \lambda t + \frac{1}{8} \lambda^2 \pi^2 K$. Logo, segue a cota que queremos mostrar. \square

Apêndice D

Teorema de Bernstein

Teorema 8 (Bernstein). *Se X_1, \dots, X_n são variáveis aleatórias independentes com média comum igual a zero e $\mathbb{P}\{|X_i| \leq c\} = 1$ para algum $c > 0$, então para algum $t > 0$,*

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i\right| > t\right\} \leq 2 \exp\left(\frac{-t^2}{2(\sigma^2 + ct/3)}\right),$$

onde $\sigma^2 = \sum_{i=1}^n \text{Var}(X_i)$.

Para demonstrar o teorema acima, vamos precisar dos seguintes resultados.

Lema 18. *Suponha que $|X| \leq c$ e $\mathbb{E}[X] = 0$. Então para todo $t > 0$,*

$$\mathbb{E}\left[e^{tX}\right] \leq \exp\left(t^2\sigma^2\left(\frac{e^{tc} - 1 - tc}{(tc)^2}\right)\right),$$

onde $\sigma^2 = \text{Var}(X)$.

Demonstração. Seja $F = \sum_{r=2}^{\infty} \frac{t^{r-2}\mathbb{E}[X^r]}{r!\sigma^2}$. Assim, pela série de Taylor temos que

$$\mathbb{E}[e^{tX}] = \mathbb{E}\left[1 + tX + t^2 \sum_{r=2}^{\infty} \frac{t^{r-2}\mathbb{E}[X^r]}{r!}\right] = 1 + t^2\sigma^2 F \leq e^{t^2\sigma^2 F}. \quad (\text{D.1})$$

Como $|X| \leq c$, logo para $r \geq 2$ temos que $\mathbb{E}[X^r] = \mathbb{E}[X^{r-2}X^2] \leq c^{r-2}\sigma^2$. Desse modo,

$$F \leq \sum_{r=2}^{\infty} \frac{t^{r-2}c^{r-2}\sigma^2}{r!\sigma^2} = \frac{1}{(tc)^2} \sum_{i=2}^{\infty} \frac{(tc)^i}{i!} = \frac{e^{tc} - 1 - tc}{(tc)^2}. \quad (\text{D.2})$$

Concluimos a prova substituindo (D.2) em (D.1). □

Lema 19. Para todo $u \geq 0$, temos que

$$(1 + u) \log(1 + u) - u \geq \frac{u^2}{2 + \frac{2u}{3}}.$$

Demonstração. Veja que $\frac{u^2}{2 + \frac{2u}{3}} = \frac{3u^2}{2(3+u)}$. Para todo $u \geq 0$, sejam $h(u) = (1 + u) \log(1 + u) - u$ e $g(u) = \frac{3u^2}{2(3+u)}$. Observe que, $h(0) = g(0) = 0$, $h'(0) = g'(0) = 0$, $h''(0) = g''(0) = 1$, $h^{(3)}(0) = g^{(3)}(0) = -1$ e

$$\frac{h^{(n)}(0)}{n!} = (-1)^{n-2} \frac{1}{n(n-1)} \geq \frac{g^{(n)}(0)}{n!} = (-1)^{n-2} \frac{3^{3-(n+1)}}{2}$$

para todo $n \geq 4$. Portanto, pela série de Taylor temos que

$$h(u) = \sum_{n=0}^{\infty} \frac{h^{(n)}(0)}{n!} u^n \geq \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} u^n = g(u),$$

para todo $u \geq 0$. □

Demonstração do Teorema 8. Veja que pelo Lema 18 temos que

$$\mathbb{E} \left[e^{tX_i} \right] \leq \exp \left(t^2 \sigma_i^2 \left(\frac{e^{tc} - 1 - tc}{(tc)^2} \right) \right), \quad (\text{D.3})$$

onde $\sigma_i^2 = \mathbb{E}[X_i^2]$. Agora, pelo truque de Chernoff temos que

$$\begin{aligned} \mathbb{P} \left\{ \sum_{i=1}^n X_i > t \right\} &= \mathbb{P} \left\{ e^{\lambda \sum_{i=1}^n X_i} > e^{\lambda t} \right\} \\ &\leq e^{-\lambda t} \mathbb{E} \left[e^{\lambda \sum_{i=1}^n X_i} \right] \\ &\stackrel{(i)}{=} e^{-\lambda t} \prod_{i=1}^n \mathbb{E} \left[e^{\lambda X_i} \right] \\ &\stackrel{(ii)}{\leq} e^{-\lambda t} \exp \left(\lambda^2 \sigma^2 \frac{e^{\lambda c} - 1 - \lambda c}{(\lambda c)^2} \right), \end{aligned}$$

onde em (i) usamos que as variáveis aleatórias X_1, \dots, X_n são independentes e em (ii) usamos (D.3). Tomando $\lambda = (1/c) \log(1 + tc/\sigma^2)$ na desigualdade acima temos que

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i > t \right\} \leq \exp \left(-\frac{\sigma^2}{c^2} h \left(\frac{tc}{\sigma^2} \right) \right),$$

onde $h(u) = (1 + u) \log(1 + u) - u$. Pelo Lema 19 temos que

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i > t \right\} \leq \exp \left(\frac{-t^2}{2(\sigma^2 + ct/3)} \right),$$

concluindo a prova do teorema. □

Bibliografia

- Addario-Berry, L., Broutin, N., Devroye, L., and Lugosi, G. (2010). On combinatorial testing problems. *The Annals of Statistics*, 38(5):3063–3092.
- Arias-Castro, E., Bubeck, S., and Lugosi, G. (2012). Detection of correlations. *The Annals of Statistics*, 40(1):412–435.
- Arias-Castro, E., Candès, E., and Durand, A. (2009). Detection of an abnormal cluster in a network. *The Bulletin of the International Statistical Association, Durban, South Africa*.
- Arias-Castro, E., Candès, E. J., Helgason, H., and Zeitouni, O. (2008). Searching for a trail of evidence in a maze. *The Annals of Statistics*, 36(4):1726–1757.
- Arias-Castro, E. and Ying, A. (2019). Detection of sparse mixtures: Higher criticism and scan statistic. *Electronic Journal of Statistics*, 13(1):208–230.
- Berman, S. M. (1962). Equally correlated random variables. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 155–156.
- Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. *Sankhyā: the indian journal of statistics*, pages 401–406.
- Brennan, S. M., Mielke, A. M., Torney, D. C., and Maccabe, A. B. (2004). Radiation detection with distributed sensor networks. *Computer*, 37(8):57–59.
- Cai, T. T. and Wu, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Transactions on Information Theory*, 60(4):2217–2232.
- Cayon, L., Jin, J., and Treaster, A. (2005). Higher criticism statistic: detecting and identifying non-gaussianity in the wmap first-year data. *Monthly Notices of the Royal Astronomical Society*, 362(3):826–832.

- Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.
- Culler, D., Estrin, D., and Srivastava, M. (2004). Guest editors' introduction: Overview of sensor networks. *Computer*, 37(8):41–49.
- Devroye, L., Györfi, L., and Lugosi, G. (1996). A probabilistic theory of pattern recognition. Technical report.
- Dobrusin, R. (1958). A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws. *Theory of Probability & Its Applications*, 3(2):161–173.
- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32(3):962–994.
- Feder, T. and Mihail, M. (1992). Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 26–38.
- Ingster, J. (1996). *On Some Problems of Hypothesis Testing Leading to Infinitely Divisible Distributions*. Weierstrass-Institut für Angewandte Analysis und Stochastik Berlin: Preprint. Weierstraß-Inst. für Angewandte Analysis und Stochastik.
- Ingster, Y. I. (1998). Minimax detection of a signal for i (n)-balls. *Mathematical Methods of Statistics*, 7(4):401–428.
- James, B. R. (1996). *Probabilidade: um curso em nível intermediário*. Number 519.2.
- Jeng, X. J., Cai, T. T., and Li, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *Journal of the American Statistical Association*, 105(491):1156–1166.
- Jin, J., Starck, J.-L., Donoho, D. L., Aghanim, N., and Forni, O. (2005). Cosmological non-gaussian signature detection: Comparing performance of different statistical tests. *EURASIP Journal on Advances in Signal Processing*, 2005(15):297184.
- Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295.

Kailath, T. and Poor, H. V. (1998). Detection of stochastic processes. *IEEE Transactions on Information Theory*, 44(6):2230–2231.

Moon, J. W. (1970). Counting labelled trees.

Talagrand, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer Monographs in Mathematics. Springer Berlin Heidelberg.