

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
Departamento de Estatística

**COMPARAÇÃO DE TESTES DE ADEQUABILIDADE DE
AJUSTE
EM DISTRIBUIÇÕES MULTINOMIAIS**

ANDRÉ JALLES MONTEIRO

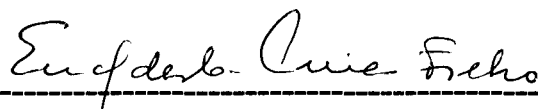
ORIENTADOR: PROF. DR. EUCLYDES CUSTÓDIO DE LIMA FILHO.

**Campinas - São Paulo
Agosto de 1995**

COMPARAÇÃO DE TESTES DE ADEQUABILIDADE DE AJUSTE
EM DISTRIBUIÇÕES MULTINOMIAIS

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. *André Jalles Monteiro* e aprovada pela Comissão Julgadora.

Campinas, 25 de agosto de 1995



Prof. Dr. Euclides Custódio de Lima Filho

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

Tese defendida e aprovada em, 25 de 08 de 1995

Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). REGINA CÉLIA CARVALHO PINTO MORAN



Prof(a). Dr(a). EUGÊNIA MARIA REGINATO CHARNET



Prof(a). Dr(a). EUCLYDES CUSTÓDIO DE LIMA FILHO

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP

764p Monteiro, André Jalles
Comparação de Testes de Adequabilidade de Ajuste em
Distribuições Multinomiais / André Jalles Monteiro.
Campinas, [SP : s.n.], 1995.

Orientador : Euclides Custódio de Lima Filho
Dissertação (mestrado) - Universidade Estadual de Campinas,
Instituto de Matemática, Estatística e Ciência da Computação.

1. Testes de hipóteses estatísticas. 2. Estatística não
paramétrica. 3. Teoria da informação. I. Lima filho, Euclides
Custódio de. II. Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Ciência da Computação. III. Título.

Homenagem Póstuma,

Quero salientar que esta dissertação não seria possível sem a primeira orientação do Prof. Dr. Hermann G. Rohrer.

Agradecimentos,

Quero agradecer a todos que direta ou indiretamente participaram na elaboração desta dissertação:

Ao Prof. Dr. Euclides Custódio de Lima Filho, pela orientação final deste trabalho, bem como ao apoio necessário.

Às Professoras: Regina Célia C. P. Moran e Eugênia M. R. Charnet, pela participação na banca examinadora, bem como pelas sugestões apresentadas.

Às instituições: CNPq e CAPES, pelo apoio financeiro.

Ao Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, pelo apoio.

Aos funcionários da UNICAMP, pelo suporte necessário.

Aos amigos e companheiros do IMECC, pela cooperação e afeto.

Dedicação,

**A minha esposa, Fabiana
e ao meu filho, Ariel.**

ÍNDICE

INTRODUÇÃO

Introdução	1
------------	---

CAPÍTULO 0

Motivação	3
-----------	---

CAPÍTULO I

1. Entropia	5
1.1. Introdução	5
1.2. Definição	6
1.3. Caso particular, $k = 2$	6
1.4. Algumas propriedades	7
1.5. Caracterização de Entropia	12
2. Divergência Direcionada	13
2.1. Introdução	13
2.2. Definição	14
2.3. Caso particular, $k = 2$	14
2.4. Algumas propriedades	15
2.5. Caracterização da Divergência Direcionada	17
3. Divergência Direcionada Generalizada	18
3.1. Introdução	18
3.2. Alguns exemplos	19
3.3. Divergência Direcionada de Kullback	20
3.4. Divergência Direcionada de Cressie e Read	22

CAPÍTULO II

1. Introdução	25
2. Teste de Hipóteses Multinomial	26
3. Breve histórico de testes Multinomiais	27
4. Teste Multinomial apresentado por Cressie e Read	29
4.1. Introdução	29
4.2. Função de Divergência	30

CAPÍTULO III

1. Análise em pequenas amostras	36
2. Aplicação da sugestão de Radlow e Alf, em testes Multinomiais	37
3. Análise de casos particulares	38
4. Procedimento para construção das regiões de rejeição	39
4.1. Teste Qui-quadrado	39
4.2. Teste como sugerido por Radlow e Alf	40
4.3. Aleatorização	40
5. Máscara de leitura dos gráficos	41
6. Hipótese nula dada por (0.1 , 0.2 , 0.7)	43
6.1. Tamanho da amostra igual a 30	43
6.2. Tamanho da amostra igual a 35	46
7. Hipótese nula dada por (0.2 , 0.3 , 0.5)	49
7.1. Tamanho da amostra igual a 20	49
7.2. Tamanho da amostra igual a 30	52
8. Hipótese nula dada por (1/3 , 1/3 , 1/3)	55
8.1. Tamanho da amostra igual a 20	55
8.2. Tamanho da amostra igual a 30	58
9. Análise dos casos particulares	61
9.1. Hipótese nula dada por (0.1 , 0.2 , 0.7)	61
9.2. Hipótese nula dada por (0.2 , 0.3 , 0.5)	63
9.3. Hipótese nula dada por (1/3 , 1/3 , 1/3)	64

CAPÍTULO IV

1. Aplicação da Teoria da Informação na Estatística	66
2. Divergência Direcionada	67
2.1. Divergência Direcionada em Testes de Hipóteses	68
3. Comparação de Testes	68
3.1. Comparação Assintótica	69
3.2. Comparação em amostras finitas	70
4. Conclusões finais	72

APÊNDICE 1

Programa para formar regiões de rejeição em testes Multinomiais aleatorizados	74
---	----

APÊNDICE 2

Programa para formar regiões de rejeição em testes Multinomiais de Cressie e Read	77
---	----

BIBLIOGRAFIA

Bibliografia	79
--------------	----

INTRODUÇÃO

Com o trabalho de Cochran, W. G. (1952) [3], temos o início de uma discussão entre qual dos testes: χ^2 , apresentado por Pearson, K. (1900) [17] e o teste do logaritmo da razão de verossimilhança, melhor se adapta em testes de adequabilidade de ajuste em distribuições multinomiais. Muitos trabalhos, posteriores ao citado, ampliaram essa discussão, onde, em alguns casos, temos inclusive novas sugestões de testes.

No trabalho de Cressie, N. A. C. e Read, T. R. C. (1984) [5], nos é apresentado uma família de estatísticas que podem ser utilizadas em testes de adequabilidade de ajuste em distribuições multinomiais. As estatísticas: χ^2 , apresentada por Pearson, K (1900) [17] e Logaritmo da Razão de Verossimilhança, apresentada por Cochran, W. G. (1952) [3] bem como algumas outras estatísticas apresentadas em outros trabalhos fazem parte dessa família de estatísticas. Nesse trabalho de Cressie, N. A. C. e Read, T. R. C. (1984) [5] temos, após análise dessa família de estatísticas, a sugestão de uma nova estatística, pertencente a essa família de estatísticas, como a que melhor se adapta para testes de adequabilidade de ajuste em distribuições multinomiais.

No presente trabalho tentamos contribuir para essa mencionada discussão, abordando alguns pontos não completamente analisados, realizando uma comparação das estatísticas pertencentes a essa família de estatísticas apresentadas em Cressie, N. A. C. e Read, T. R. C. (1984) [5] por intermédio das regiões de rejeição denotada por essas estatísticas, sugerindo a semelhança entre os testes χ^2 , apresentada por Pearson, K (1900) [17] e Logaritmo da Razão de Verossimilhança, apresentada por Cochran, W. G. (1952) [3], marco inicial da discussão, bem como muitos outros que utilizam as estatísticas da referida família de estatísticas. Não temos a pretensão de finalizar essa longa discussão, porém amplia-la com novas argumentações.

No capítulo 0, temos uma primeira abordagem do assunto que trata este trabalho, além de propiciar ao leitor a visão com que o mesmo será estudado, no desenvolver deste trabalho. Temos nesse capítulo uma introdução um pouco mais abrangente do conteúdo dos capítulos seguintes.

No capítulo I, temos uma breve revisão de elementos da Teoria da Informação, que admitimos, devido a bibliografia citada, ser o ferramental teórico inicial do trabalho de Cressie, N. A.

C. e Read, T. R. C. (1984) [5]. Apresentamos a Entropia, como definida em Shannon, E. C. (1948) [16], bem como a Divergência Direcionada, abordada no trabalho de Kullback, S. (1959) [7], e posteriormente generalizada por diversos autores como: Rényi, Rathie e Kannappan, Cressie e Read.

No capítulo II, fazemos a ligação da Teoria da Informação com testes de adequabilidade de ajuste em distribuições multinomiais. Temos também um breve histórico do desenvolvimento de testes de adequabilidade de ajuste em distribuições multinomiais, bem como apresentamos uma síntese do trabalho de Cressie, N. A. C. e Read, T. R. C. (1984) [5].

No capítulo III, apresentamos uma simulação de alguns casos particulares para alguns testes utilizando as estatísticas definidas na família de estatísticas apresentadas em Cressie, N. A. C. e Read, T. R. C. (1984) [5], fazendo em seguida análises e comentários dos resultados obtidos, que mesmo não expressando todos os possíveis casos, nos revela uma amostra do que acontece em testes de adequabilidade de ajuste em distribuições multinomiais utilizando essa família de estatísticas.

No capítulo IV, apresentamos as conclusões desse trabalho, partindo-se desde a relação de Teoria da Informação e testes de adequabilidade de ajuste em distribuições multinomiais, até as análises, em termos mais gerais, da simulação apresentada no capítulo anterior, incluindo-se a essa análise comentários também da análise assintótica dos testes da mencionada família de estatísticas.

CAPÍTULO 0

MOTIVAÇÃO

Na Inferência Estatística nos deparamos com um conjunto de dados, que assumimos ser oriundo da realização de alguma variável aleatória, que admitimos pertencer a um certo modelo cuja função de probabilidade ou densidade de probabilidade é parametrizável e pretendemos, através de algum mecanismo, obter algum conhecimento desses parâmetros por intermédio desse respectivo conjunto de dados.

Em alguns casos podemos ter interesse em um modelo que apresente variáveis aleatórias discretas ou discretizáveis, assumindo somente um número finito de valores distintos, onde o conjunto de dados, ao qual estamos associando este modelo é obtido em realizações independentes de uma variável aleatória pertencente a este modelo. Nos interessa saber a possível função de probabilidade da respectiva variável aleatória. Tratamos no caso de observações independentes de variáveis aleatórias que assumem somente um número finito de valores distintos, pois temos a pretensão de analisarmos somente o caso de distribuições multinomiais.

Um mecanismo possível para a associação desse conjunto de dados à uma variável aleatória pode ser a escolha, dentro de opções pré estabelecidas, de uma ou um grupo de variáveis aleatórias que possuam uma maior plausibilidade com os respectivos dados, ou melhor, dentre duas opções de variáveis aleatórias ou grupos de variáveis aleatórias selecionamos o que melhor se adequa na possibilidade de ser a responsável pela geração do conjunto de dados em questão.

Porém, podemos ter interesse em uma certa variável aleatória particular, ou seja, temos a pretensão de explicar o surgimento de um certo conjunto de dados através de uma particular variável aleatória do modelo acima descrito, ou seja, o conjunto de dados é oriundo de observações independentes de uma variável aleatória que assume somente um número finito de valores distintos.

Podemos então pensar na possibilidade de medirmos o grau de plausibilidade existente entre essa variável aleatória de interesse e os respectivos dados apresentados, onde essa medida de plausibilidade deve levar em consideração as possíveis variáveis aleatórias alternativas, para a explicação do surgimento do conjunto de dados em questão.

Ao atribuímos o surgimento de um conjunto de dados à realizações independentes de uma certa variável aleatória particular, não podemos nos contrapor a possibilidade de tolerarmos, como geradora do conjunto de dados em questão, variáveis aleatórias, cujas funções de probabilidade tenham valores muito próximos à função de probabilidade dessa certa variável aleatória.

Uma primeira abordagem desse tema pode ser visto em Pearson, K. (1900) [17] quando na generalização da estatística χ^2 , mostrando sua convergência para a distribuição Qui-quadrado, sendo posteriormente confrontado em Cochran, W. G. (1952) [3] com uma nova estatística, a razão do logaritmo da verossimilhança, mostrando este que esta nova estatística, sob alguns aspectos, possuía propriedades mais atraentes do que aquela, sendo então melhor, sob esses aspectos, para resolver problemas como os acima citados. Posteriormente a W. G. Cochran, muitos autores defenderam ora esta ora aquela estatística como a mais adequada na resolução desses problemas, tendo também alguns autores novas sugestões de estatísticas. Dentro desse assunto N. A. C. Cressie e T. R. C. Read construíram uma família de estatísticas, que continham como casos particulares algumas dessas estatísticas sugeridas.

O trabalho de N. A. C. Cressie e T. R. C. Read nos dá boas chances de questionarmos e compararmos essas estatísticas, porém este trabalho tem como início a apresentação de uma família de estatísticas que não sabemos ao certo de onde surgiram. Partindo-se da teoria da informação, bem como uma ligação desta com a inferência estatística encontramos uma chave para essa explicação, deixando-nos mais seguros para a utilização deste trabalho na comparação dessas estatísticas.

Esquecendo-se um pouco da idéia de poder de testes de hipóteses, já que para esses casos não possuímos um teste mais poderoso em todas as opções de hipóteses alternativas, buscamos na divisão do espaço amostral em região de maior plausibilidade para a hipótese de interesse e região com menor plausibilidade, medida essa plausibilidade por intermédio dessas estatísticas apresentadas no trabalho de N. A. C. Cressie e T. R. C. Read, mostrando-se as diferenças e/ou semelhanças entre essas estatísticas por intermédio das diferenças dessas regiões, medindo-se assim a afinidade entre essas estatísticas.

CAPÍTULO I

ENTROPIA E DIVERGÊNCIA DIRECIONADA

1. Entropia

1.1. Introdução

A Entropia, como definida em Shannon, E. C. (1948) [16], pode ser vista como uma medida, do grau de flutuação do ponto que será obtido na realização de uma variável aleatória discreta. Essa medida está associada à função de probabilidade de uma variável aleatória discreta, e é encontrada partindo-se de propriedades requeridas para uma quantificação dessa natureza. Outros autores como Ash, R. (1965) [1] e Rényi, A. (1970) [15] chegam a mesma medida, partindo de propriedades diferentes, porém algumas dessas propriedades repetem-se nos três autores, bem como em muitos outros.

A passagem para o caso contínuo, ou seja, quando a variável aleatória em questão é absolutamente contínua, é quase que imediato, porém nesse trabalho enfatizamos somente o caso discreto.

Seja S_k o conjunto de todas as variáveis aleatórias discretas que assumem, no máximo, k valores distintos e P_k a família de distribuições de probabilidades dessas variáveis aleatórias, onde R representa o conjunto dos números reais, temos:

$$S_k = \left\{ \hat{x} = (x_1, x_2, \dots, x_k) : x_i \in \mathcal{O}, \forall i = 1, 2, \dots, k \text{ e } \sum_{i=1}^k P[X = x_i] = 1 \right\}$$

$$P_k = \left\{ \hat{p} = (p_1, p_2, \dots, p_k) : p_i \geq 0, \forall i = 1, 2, \dots, k \text{ e } \sum_{i=1}^k p_i = 1 \right\}$$

1.2. Definição

A Entropia da função de probabilidade de uma variável aleatória discreta, com função de probabilidades em P_k , definida em Shannon, E. C. (1948) [16], é dada por:

$$H_k(p_1, p_2, \dots, p_k) = - \sum_{i=1}^k p_i \text{Log}_2 p_i$$

assumindo, por convenção que $0 \cdot \text{Log}_2 0 = 0$, pois: $\lim_{p \rightarrow 0^+} p \cdot \text{Log}_2 p = 0$

1.3. Caso particular, $k = 2$.

Um caso particular, quando $k = 2$, é de grande importância, possuindo algumas propriedades interessantes, como apresentado em Shannon, E. C. (1948) [16].

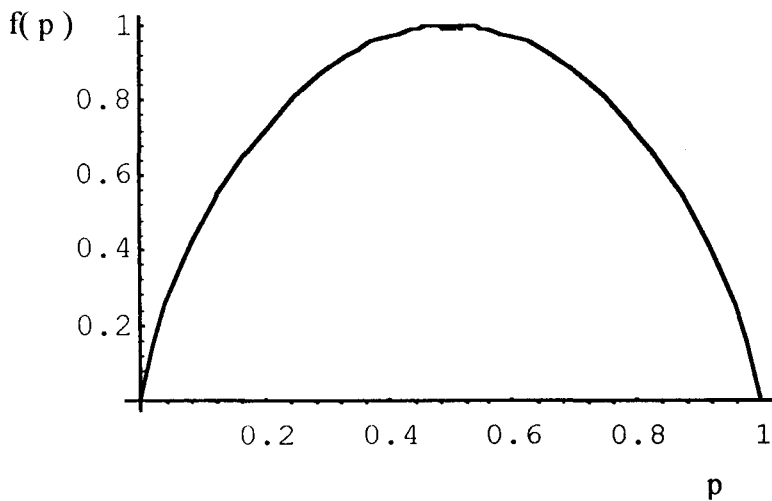
$$H_2(p, 1-p) = -p \cdot \text{Log}_2(p) - (1-p) \cdot \text{Log}_2(1-p)$$

observando-se em função de p , temos:

$$f(p) = -p \cdot \text{Log}_2(p) - (1-p) \cdot \text{Log}_2(1-p)$$

apresentando o gráfico que segue:

gráfico 1.1: $f(p) = -p \cdot \text{Log}_2(p) - (1-p) \cdot \text{Log}_2(1-p)$, $p \in [0, 1]$.



1.4. Algumas propriedades de Entropia

1.4.1. Não negatividade

$$H_k(p_1, p_2, \dots, p_k) \geq 0, \forall (p_1, p_2, \dots, p_k) \in P_k.$$

prova:

$$H_k(p_1, p_2, \dots, p_k) = -\sum_i p_i \cdot \text{Log}_2 p_i, \quad i=1, 2, \dots, k$$

i) se $p_i = 0$, $p_i \cdot \text{Log}_2 p_i = 0$, por convenção.

ii) se $0 < p_i \leq 1$, $\text{Log}_2 p_i \leq 0 \Rightarrow p_i \cdot \text{Log}_2 p_i \leq 0$

por (i) e (ii) $\sum_i p_i \cdot \text{Log}_2 p_i \leq 0 \Rightarrow -\sum_i p_i \cdot \text{Log}_2 p_i \geq 0$ ■

1.4.2. Aditividade

Sejam $(p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_h) \in P_h$, onde $p_i > 0, \forall i=1, 2, \dots, k$ e $q_i > 0, \forall i=1, 2, \dots, h$.

$$H_{k,h}(p_1 \cdot q_1, \dots, p_1 \cdot q_h, p_2 \cdot q_1, \dots, p_k \cdot q_h) = H_k(p_1, p_2, \dots, p_k) + H_h(q_1, q_2, \dots, q_h)$$

prova:

$$\sum_{i=1}^k \sum_{j=1}^h p_i \cdot q_j = \sum_{i=1}^k p_i \cdot \sum_{j=1}^h q_j = \sum_{i=1}^k p_i = 1$$

$$H_{k,h}(p_1 \cdot q_1, \dots, p_1 \cdot q_h, p_2 \cdot q_1, \dots, p_k \cdot q_h) = -\sum_{i=1}^k \sum_{j=1}^h p_i \cdot q_j \cdot \text{Log}_2 (p_i \cdot q_j) =$$

$$\begin{aligned}
&= -\sum_{i=1}^k \sum_{j=1}^h p_i \cdot q_j \cdot (\text{Log}_2 p_i + \text{Log}_2 q_j) = -\sum_{i=1}^k \sum_{j=1}^h p_i \cdot q_j \cdot \text{Log}_2 p_i - \sum_{i=1}^k \sum_{j=1}^h p_i \cdot q_j \cdot \text{Log}_2 q_j = \\
&= -\sum_{i=1}^k p_i \cdot \text{Log}_2 p_i \cdot \sum_{j=1}^h q_j - \sum_{j=1}^h q_j \cdot \text{Log}_2 q_j \cdot \sum_{i=1}^k p_i = \left(\text{como, } \sum_{i=1}^k p_i = \sum_{j=1}^h q_j = 1 \right) = \\
&= -\sum_{i=1}^k p_i \cdot \text{Log}_2 p_i - \sum_{j=1}^h q_j \cdot \text{Log}_2 q_j = H_k(p_1, p_2, \dots, p_k) + H_h(q_1, q_2, \dots, q_h) \bullet
\end{aligned}$$

1.4.3. Máximo em equiprobabilidade

$$\text{Sup } H_k(p_1, p_2, \dots, p_k) = H_k(1/k, 1/k, \dots, 1/k), (p_1, p_2, \dots, p_k) \in P_k.$$

1.4.3.1. Lema: $\text{Log}_e(x) \leq x - 1, \forall x > 0$.

prova:

seja $f(x) = x - 1 - \text{Log}_e(x), x > 0$.

$$i) \frac{\partial f(x)}{\partial x} = 1 - \frac{1}{x} \quad \frac{\partial f(x)}{\partial x} \Big|_{x=1} = 0.$$

$$\text{ii) } \frac{\partial^2 f(x)}{\partial x^2} = \frac{1}{x^2} > 0, \forall x > 0, \text{ sendo então } f(x) \text{ uma função côncava.}$$

por (i) e (ii) podemos dizer que o ponto $x = 1$ é um mínimo global, como $f(1) = 0$, podemos dizer que a função $f(x)$ é não negativa, ou seja, $f(x) \geq 0, \forall x > 0 \Rightarrow x - 1 - \text{Log}_e(x) \geq 0 \Rightarrow \text{Log}_e(x) \leq x - 1, \forall x > 0.$

1.4.3.2. Lema: Sejam: $(p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_k) \in P_k$, com $p_i > 0$ e $q_i > 0, \forall i=1, 2, \dots, k.$
Então: $\sum_i p_i \cdot \text{Log}_e p_i \geq \sum_i p_i \cdot \text{Log}_e q_i, i=1, 2, \dots, k.$

prova:

usando o lema 1.4.3.1. temos que: $\text{Log}_e(x) \leq x - 1, \forall x > 0,$ fazendo-se $x = p_i/q_i,$ temos que:

$$\begin{aligned} \text{Log}_e(q_i/p_i) \leq q_i/p_i - 1 &\Rightarrow p_i \cdot \text{Log}_e(q_i/p_i) \leq q_i - p_i \Rightarrow \sum_{i=1}^k p_i \cdot \text{Log}_e(q_i/p_i) \leq \sum_{i=1}^k (p_i - q_i) \Rightarrow \\ \Rightarrow \sum_{i=1}^k p_i \cdot (\text{Log}_e q_i - \text{Log}_e p_i) \leq 0 &\Rightarrow \sum_{i=1}^k p_i \cdot \text{Log}_e p_i \geq \sum_{i=1}^k p_i \cdot \text{Log}_e q_i \bullet \end{aligned}$$

prova de 1.4.3:

usando o lema 1.4.3.2. temos que: $\sum_i p_i \cdot \text{Log}_e p_i \geq \sum_i p_i \cdot \text{Log}_e q_i, i = 1, 2, \dots, k, \forall (p_1, p_2, \dots, p_k), (q_1, q_2, \dots, q_k) \in P_k,$ com $p_i > 0$ e $q_i > 0, \forall i=1, 2, \dots, k.$ Fazendo-se $q_i = 1/k, \forall i=1, 2, \dots, k,$ temos que:

$$\sum_{i=1}^k p_i \cdot \text{Log}_e p_i \geq \sum_{i=1}^k p_i \cdot \text{Log}_e (1/k) = \text{Log}_e (1/k) \cdot \sum_{i=1}^k p_i = \text{Log}_e (1/k) = \sum_{i=1}^k (1/k) \cdot \text{Log}_e (1/k)$$

$$\Rightarrow \frac{-1}{\text{Log}_e 2} \cdot \sum_{i=1}^k p_i \cdot \text{Log}_e p_i \leq \frac{-1}{\text{Log}_e 2} \cdot \sum_{i=1}^k (1/k) \cdot \text{Log}_e (1/k) \Rightarrow -\sum_{i=1}^k p_i \cdot \text{Log}_2 p_i \leq -\sum_{i=1}^k (1/k) \cdot \text{Log}_2 (1/k) \bullet$$

1.4.4. Sub-aditividade

Para $\{p_{i,j}\}_{i=1,2,\dots,k \text{ e } j=1,2,\dots,h}$, com $p_{i,j} \in P_{k,h}$, temos:

$$H_{k,h}(p_{1,1}, \dots, p_{1,h}, p_{2,1}, \dots, p_{k,h}) \leq H_k(\sum_j p_{1,j}, \sum_j p_{2,j}, \dots, \sum_j p_{k,j}) + H_h(\sum_i p_{i,1}, \sum_i p_{i,2}, \dots, \sum_i p_{i,h}).$$

prova:

sejam: $p_i = \sum_j p_{i,j}$ e $p_j = \sum_i p_{i,j}$.

$$\begin{aligned} H_{k,h}(p_{1,1}, \dots, p_{1,h}, p_{2,1}, \dots, p_{k,h}) &= -\sum_{i=1}^k \sum_{j=1}^h p_{i,j} \cdot \text{Log}_2 p_{i,j} \leq -\sum_{i=1}^k \sum_{j=1}^h p_{i,j} \cdot \text{Log}_2 (p_i \cdot p_j) = \\ &= -\sum_{i=1}^k \sum_{j=1}^h p_{i,j} \cdot (\text{Log}_2 p_i + \text{Log}_2 p_j) = -\sum_{i=1}^k \sum_{j=1}^h p_{i,j} \cdot \text{Log}_2 p_i - \sum_{i=1}^k \sum_{j=1}^h p_{i,j} \cdot \text{Log}_2 p_j = \\ &= -\sum_{i=1}^k p_i \cdot \text{Log}_2 p_i - \sum_{j=1}^h p_j \cdot \text{Log}_2 p_j = H_k(p_1, p_2, \dots, p_k) + H_h(p_1, p_2, \dots, p_h) \bullet \end{aligned}$$

1.5. Caracterização de Entropia

A medida de Entropia como sugerida por Shannon, E. C. (1948) [16] pode ser encontrada partindo-se de certas propriedades requeridas para essa medida de incerteza. Vários autores, posteriormente a Shannon, sugeriram diferentes propriedades. Em Kannappan, P. (1972) [6] essas propriedades ficam parcialmente reduzidas à uma equação funcional, nesse trabalho podemos ver a medida de Entropia, como definida por Shannon, por intermédio da solução da seguinte equação funcional:

$\sum_i \sum_j f(x_i \cdot y_j) = \sum_i f(x_i) + \sum_j f(y_j)$, com $i=1,2,\dots,k$ e $j=1,2,\dots,h$; onde k e h são números inteiros maiores que um, $x_i \geq 0$, $i=1,2,\dots,k$ e $y_j \geq 0$, $j=1,2,\dots,h$.

Sendo essa solução dada por: $f(x) = A \cdot x \cdot \text{Log}_e(x)$, de tal modo que A é uma constante, podendo-se assim obtermos o logaritmo, relacionado pela função, em qualquer base desejada. Fazendo-se $A = -1/\text{Log}_e 2$, através de alguma outra propriedade específica, chegamos à fórmula de Shannon.

Outros autores, como Mathai, A. M. e Rathie, P. N. (1975) [10] chegam à medida de Entropia através das seguintes propriedades:

i- Continuidade: $H_k(p_1, p_2, \dots, p_k)$ é contínua em relação a (p_1, p_2, \dots, p_k) .

ii- Monotonicidade: $H_k(1/k, 1/k, \dots, 1/k) \leq H_{k+1}(1/(k+1), 1/(k+1), \dots, 1/(k+1))$.

iii- Recursividade: $H_k(p_1, p_2, \dots, p_{k-1}, p_k \cdot q_1, p_k \cdot q_2, \dots, p_k \cdot q_h) = H_k(p_1, p_2, \dots, p_k) + p_k \cdot H_h(q_1, q_2, \dots, q_h)$,

onde: $(p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_h) \in P_h$.

iv- Normalização: $H_2(0.5, 0.5) = 1$.

v- Simetria: $H_k(p_1, p_2, \dots, p_k) = H_k(p_{i_1}, p_{i_2}, \dots, p_{i_k})$, onde $\{i_1, i_2, \dots, i_k\}$ é uma permutação de $\{1, 2, \dots, k\}$.

2. Divergência Direcionada

2.1. Introdução

Partindo-se do conceito de Entropia, definido por Shannon E. C. (1948) [16], podemos mensurar o grau de incerteza gerado pela função de probabilidade de uma variável aleatória discreta, ou categorizada em k células mutuamente exclusivas, onde essa medida é máxima quando essas células são igualmente prováveis, ou seja, no caso equiprovável e mínima (zero) somente no caso de haver total concentração de probabilidade em uma única célula, ou seja, uma categoria com probabilidade um, com as restantes somando probabilidade zero.

Podemos ampliar o conceito de medida de entropia para uma medida de divergência entre a distribuição de probabilidade de uma variável aleatória categorizada em k células mutuamente exclusivas e a correspondente equiprovável. Uma generalização natural será a de mensurar a divergência entre distribuições de probabilidade de duas variáveis aleatórias quaisquer, dessa mesma natureza. Esse termo foi primeiramente abordado por Kullback, S. e Leibler, R. A. (1951) [8] e novamente mencionado em Kullback, S. (1959) [7]. Posteriormente alguns outros autores como Rényi, Rathie e Kannappan, Cressie e Read, definiram outras medidas de divergências entre distribuições de probabilidades de variáveis aleatórias discretas.

2.2. Definição

Sejam: $(p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_k) \in P_k$, com $q_i > 0, \forall i=1, 2, \dots, k$. Kullback, S. (1959) [7], definiu como divergência direcionada entre as distribuições de probabilidades das variáveis aleatórias discretas com distribuições de probabilidades acima mencionadas, como:

$$K(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \sum_{i=1}^k p_i \cdot \text{Log}_2(p_i/q_i)$$

O termo divergência direcionada indica que essa medida não deve ser observada como uma medida de divergência propriamente dita, pois não possui necessariamente a propriedade de simetria, geralmente $K(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) \neq K(q_1, q_2, \dots, q_k : p_1, p_2, \dots, p_k)$, tendo então um componente direcional, ou seja, mede a divergência entre a distribuição de probabilidades (p_1, p_2, \dots, p_k) de uma variável aleatória em relação a uma certa distribuição de probabilidades (q_1, q_2, \dots, q_k) de uma outra variável aleatória.

2.3. Caso particular, $k = 2$.

Podemos ter uma certa idéia do que mede a divergência direcionada através da visualização do caso particular onde $k = 2$. Temos que:

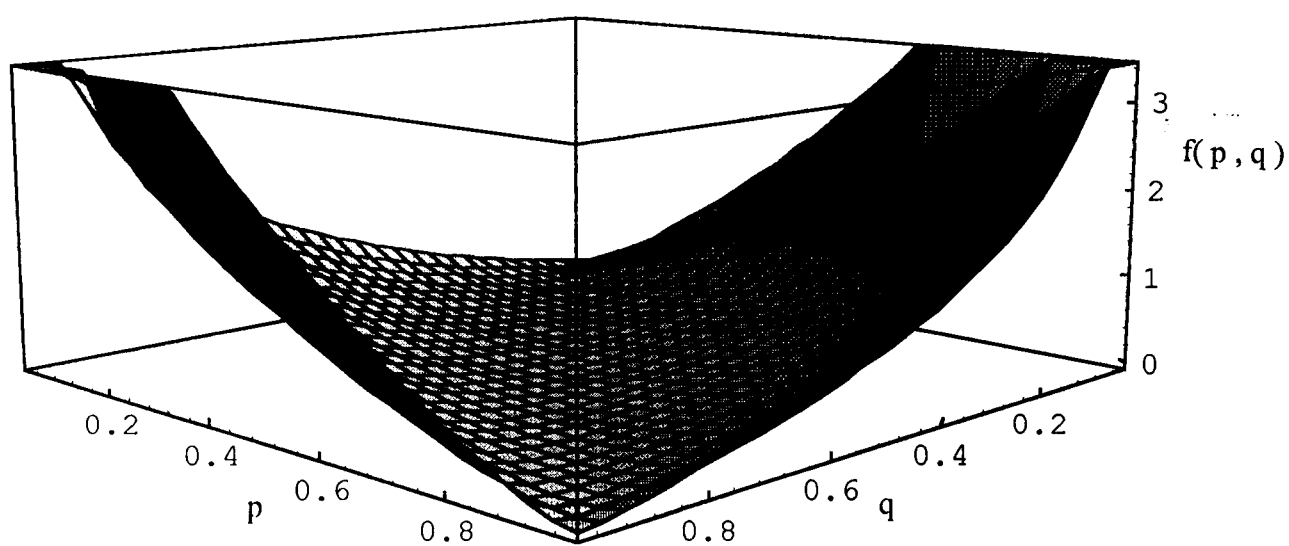
$$K(p, 1-p : q, 1-q) = p \cdot \text{Log}_2(p/q) + (1-p) \cdot \text{Log}_2((1-p)/(1-q))$$

observando-se em função de p e q , temos:

$$f(p, q) = p \cdot \text{Log}_2(p/q) + (1-p) \cdot \text{Log}_2((1-p)/(1-q)).$$

apresentando o seguinte gráfico.

gráfico 1.2: $f(p, q) = p \cdot \text{Log}_2(p/q) + (1-p) \cdot \text{Log}_2((1-p)/(1-q))$.



2.4. Algumas propriedades de divergência direcionada

2.4.1. Não negatividade

$$K(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) \geq 0, \forall (p_1, p_2, \dots, p_k) \in P_k \text{ e } (q_1, q_2, \dots, q_k) \in P_k, \text{ com } q_i > 0, \forall i=1, 2, \dots, k.$$

prova:

$$K(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \sum_{i=1}^k p_i \cdot \text{Log}_2(p_i/q_i) = \sum_{i=1}^k p_i \cdot (\text{Log}_2 p_i - \text{Log}_2 q_i) =$$

$$= \sum_{i=1}^k p_i \cdot \text{Log}_2 p_i - \sum_{i=1}^k p_i \cdot \text{Log}_2 q_i \geq 0 \text{ (lema 1.4.3.2.)} \bullet$$

2.4.2. Aditividade

Sejam: $(p_1, p_2, \dots, p_k) \in P_k$, $(q_1, q_2, \dots, q_h) \in P_h$, $(s_1, s_2, \dots, s_k) \in P_k$ e $(r_1, r_2, \dots, r_h) \in P_h$,
com $s_i > 0$, $\forall i=1, 2, \dots, k$ e $r_i > 0$, $\forall i=1, 2, \dots, h$.

$$K(p_1 q_1, \dots, p_1 q_h, p_2 q_1, \dots, p_k q_h; s_1 r_1, \dots, s_1 r_h, s_2 r_1, \dots, s_k r_h) =$$

$$= K(p_1, p_2, \dots, p_k; s_1, s_2, \dots, s_k) + K(q_1, q_2, \dots, q_h; r_1, r_2, \dots, r_h).$$

prova:

$$K(p_1 q_1, \dots, p_1 q_h, p_2 q_1, \dots, p_k q_h; s_1 r_1, \dots, s_1 r_h, s_2 r_1, \dots, s_k r_h) = \sum_{i=1}^k \sum_{j=1}^h p_i q_j \text{Log}_2 \left(\frac{p_i q_j}{s_i r_j} \right) =$$

$$= \sum_{i=1}^k \sum_{j=1}^h p_i q_j \left(\text{Log}_2 \left(\frac{p_i}{s_i} \right) + \text{Log}_2 \left(\frac{q_j}{r_j} \right) \right) = \sum_{i=1}^k \sum_{j=1}^h p_i q_j \text{Log}_2 \left(\frac{p_i}{s_i} \right) + \sum_{i=1}^k \sum_{j=1}^h p_i q_j \text{Log}_2 \left(\frac{q_j}{r_j} \right) =$$

$$\begin{aligned}
&= \sum_{i=1}^k p_i \text{Log}_2 \left(\frac{p_i}{s_i} \right) \sum_{j=1}^h q_j + \sum_{j=1}^h q_j \text{Log}_2 \left(\frac{q_j}{r_j} \right) \sum_{i=1}^k p_i = \sum_{i=1}^k p_i \text{Log}_2 \left(\frac{p_i}{s_i} \right) + \sum_{j=1}^h q_j \text{Log}_2 \left(\frac{q_j}{r_j} \right) = \\
&= K(p_1, p_2, \dots, p_k; s_1, s_2, \dots, s_k) + K(q_1, q_2, \dots, q_h; r_1, r_2, \dots, r_h) \bullet
\end{aligned}$$

2.5. Caracterização da divergência direcionada

Assim como na Entropia de Shannon, podemos obter a divergência direcionada entre distribuições de probabilidade de variáveis aleatórias discretas, através de propriedades requeridas para uma medida dessa natureza. Em kannappan, P. (1972) [6] essas propriedades ficam reduzidas à resolução de uma equação funcional, com a adição de outras duas propriedades específicas, são elas:

$$\sum_{i=1}^k \sum_{j=1}^h f(x_i \cdot y_j, u_i \cdot v_j) = \sum_{i=1}^k f(x_i, u_i) + \sum_{j=1}^h f(y_j, v_j), \text{ onde: } x_i \geq 0, y_j \geq 0, u_i > 0 \text{ e } v_j > 0,$$

$$\forall i = 1, 2, \dots, k \text{ e } j = 1, 2, \dots, h. \text{ com } \sum_{i=1}^k x_i = \sum_{j=1}^h y_j = \sum_{i=1}^k u_i = \sum_{j=1}^h v_j = 1.$$

de tal modo que:

$$\text{i) } f(1, 0.5) = 1$$

$$\text{ii) } f(0.5, 0.5) = 0$$

sendo a solução única encontrada para f , dada por: $f(x, y) = x \cdot \log_2(x/y)$, com $x \in [0, 1]$ e $y \in (0, 1]$.

Em um outro trabalho, Mathai, A. M. e Rathie, P. N. (1975) [10] chegam à divergência direcionada, definida de maneira única, através de certas propriedades, sejam elas:

i) Recursividade: $K(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = K(p_1 + p_2, p_3, \dots, p_k : q_1 + q_2, q_3, \dots, q_k) +$

$+ (p_1 + p_2) \cdot K(p_1 / (p_1 + p_2), p_2 / (p_1 + p_2) : q_1 / (q_1 + q_2), q_2 / (q_1 + q_2))$, com: $p_1 + p_2 > 0$ e $q_1 + q_2 > 0$.

ii) Simetria: $K(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = K(p_{i_1}, p_{i_2}, \dots, p_{i_k} : q_{i_1}, q_{i_2}, \dots, q_{i_k})$, onde $\{ i_1, i_2, \dots, i_k \}$ é uma permutação de $\{ 1, 2, \dots, k \}$.

iii) Derivável: Seja $f(p, q) = K(p, 1-p : q, 1-q)$, para $p \in (0, 1)$ e $q \in (0, 1)$. $f(p, q)$ possui derivadas parciais de primeira ordem contínua, com respeito a qualquer das variáveis.

iv) Normalização: $K(2/3, 1/3 : 1/3, 2/3) = 1/3$

v) Nulidade: $K(p, 1-p : p, 1-p) = 0, \forall p \in (0, 1)$.

3. Divergência direcionada generalizada

3.1. Introdução

Alguns autores, tomando como base a divergência direcionada definida em Kullback, S. (1959) [7], construíram divergências direcionadas mais gerais, passando a ser aquela, casos particulares destas.

Uma dessas generalizações pode ser visto em Rényi, A. (1970) [15], outra generalização foi apresentada em Rathie P. N. e Kannappan, P. (1972) [13]. Tomando, possivelmente, como base essa última Read, T. R. C. e Cressie, N. A. C. (1988) [14] sugeriram uma outra generalização.

Assim como em Kullback, S. (1959) [7] essas divergências direcionadas generalizadas são obtidas através de propriedades requeridas para medidas dessa natureza.

3.2. Alguns exemplos de divergências direcionadas generalizadas.

Sejam: $(p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_k) \in P_k$, com $q_i > 0$, $\forall i=1, 2, \dots, k$. Assumindo, por convenção, que: $0 \cdot \text{Log}_2 0 = 0$.

3.2.1. Divergência direcionada de ordem α de Rényi.

$$I'_{k,\alpha} (p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \frac{1}{\alpha - 1} \text{Log}_2 \left(\sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha}) \right), \alpha \neq 1.$$

3.2.2. Divergência direcionada de ordem α de Rathie e Kannappan

$$I''_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \frac{1}{2^{\alpha-1} - 1} \left(\sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha} - 1) \right), \alpha \neq 1.$$

Obs. para ambos os casos $I'_{k,1}$ e $I''_{k,1}$ são definidas, por continuidade, através dos limites quando $\alpha \rightarrow 1$.

3.3. Divergência direcionada de Kullback, vista como um caso particular

Para cada um dos casos, de divergências direcionadas generalizadas de ordem α , apresentados anteriormente podemos obter a divergência direcionada de Kullback como um caso particular, ou seja, para um valor particular de α ($\alpha = 1$).

prova:

$$I'_{k,1}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow 1} I'_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \text{Log}_2 \sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha}) =$$

$$\lim_{\alpha \rightarrow 1} \frac{1}{(\alpha - 1) \text{Log}_e 2} \text{Log}_e \sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha}) = (\text{derivando numerador e denominador}) =$$

$$= \lim_{\alpha \rightarrow 1} \frac{1}{\text{Log}_e 2} \frac{\sum_{i=1}^k \left(\text{Log}_e(p_i) \cdot (p_i^\alpha \cdot q_i^{1-\alpha}) - \text{Log}_e(q_i) \cdot (p_i^\alpha \cdot q_i^{1-\alpha}) \right)}{\sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha})} = \frac{1}{\text{Log}_e 2} \sum_{i=1}^k (p_i \cdot (\text{Log}_e p_i - \text{Log}_e q_i)) =$$

$$= \sum_{i=1}^k \left(p_i \cdot \left(\frac{\text{Log}_e p_i}{\text{Log}_e 2} - \frac{\text{Log}_e q_i}{\text{Log}_e 2} \right) \right) = \sum_{i=1}^k (p_i \cdot (\text{Log}_2 p_i - \text{Log}_2 q_i)) = \sum_{i=1}^k \left(p_i \cdot \text{Log}_2 \left(\frac{p_i}{q_i} \right) \right) \bullet$$

$$I''_{k,1}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow 1} I''_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) =$$

$$= \lim_{\alpha \rightarrow 1} \frac{1}{2^{\alpha-1} - 1} \left(\sum_{i=1}^k (p_i^\alpha \cdot q_i^{1-\alpha} - 1) \right) = (\text{derivando numerador e denominador}) =$$

$$= \lim_{\alpha \rightarrow 1} \frac{1}{2^{\alpha-1} \cdot \text{Log}_e 2} \left(\sum_{i=1}^k \left(\text{Log}_e p_i \cdot (p_i^\alpha \cdot q_i^{1-\alpha}) - \text{Log}_e q_i (p_i^\alpha \cdot q_i^{1-\alpha} - 1) \right) \right) =$$

$$= \frac{1}{\text{Log}_e 2} \left(\sum_{i=1}^k (p_i \cdot (\text{Log}_e p_i - \text{Log}_e q_i)) \right) = \sum_{i=1}^k \left(p_i \cdot \left(\frac{\text{Log}_e p_i}{\text{Log}_e 2} - \frac{\text{Log}_e q_i}{\text{Log}_e 2} \right) \right) = \sum_{i=1}^k (p_i \cdot (\text{Log}_2 p_i - \text{Log}_2 q_i)) =$$

$$= \sum_{i=1}^k \left(p_i \cdot \text{Log}_2 \left(\frac{p_i}{q_i} \right) \right) \bullet$$

3.4. Divergência direcionada de ordem α de Cressie e Read

Partindo-se da divergência direcionada de ordem α proposta por Rathie e Kannappan, com algumas alterações, evitando a desagradável situação de $I''_{k,0}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = 0$, $\forall (p_1, p_2, \dots, p_k) \in P_k$ e $(q_1, q_2, \dots, q_k) \in P_k$, com $q_i > 0$, $\forall i=1, 2, \dots, k$. Acontecendo de modo análogo à $I'_{k,0}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = 0$, ou seja, em ambos os casos a divergência direcionada generalizada entre duas quaisquer distribuições de probabilidade é zero. Podemos chegar a uma outra divergência direcionada generalizada, como definida em Cressie, N. A. C. e Read, T. R. C. (1984) [C2], temos:

$$I_{k,\alpha}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \frac{1}{\alpha \cdot (\alpha + 1)} \left(\sum_{i=1}^k \left(p_i \cdot \left(\frac{p_i}{q_i} \right)^\alpha - 1 \right) \right), \quad \alpha \notin \{-1, 0\}$$

Definindo-se, por continuidade, através do limite nos casos $\alpha = -1$ e $\alpha = 0$, obtemos:

$$I_{k,-1}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow -1} I_{k,\alpha}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \sum_{i=1}^k \left(q_i \cdot \text{Log}_e \left(\frac{q_i}{p_i} \right) \right)$$

$$I_{k,0}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow 0} I_{k,\alpha}(p_1, p_2, \dots, p_k : q_1, q_2, \dots, q_k) = \sum_{i=1}^k \left(p_i \cdot \text{Log}_e \left(\frac{p_i}{q_i} \right) \right)$$

prova:

$$I_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow a} \frac{1}{\alpha \cdot (\alpha + 1)} \left(\sum_{i=1}^k \left(p_i \cdot \left(\frac{p_i}{q_i} \right)^\alpha - 1 \right) \right), \quad a \in \{-1, 0\}$$

(derivando numerador e denominador)
$$\lim_{\alpha \rightarrow a} \frac{1}{2\alpha + 1} \left(\sum_{i=1}^k \left(p_i \cdot \text{Log}_e \left(\frac{p_i}{q_i} \right) \cdot \left(\frac{p_i}{q_i} \right)^\alpha \right) \right)$$

$$I_{k,-1}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow -1} \frac{1}{2\alpha + 1} \left(\sum_{i=1}^k \left(p_i \cdot \text{Log}_e \left(\frac{p_i}{q_i} \right) \cdot \left(\frac{p_i}{q_i} \right)^\alpha \right) \right) = \sum_{i=1}^k \left(q_i \cdot \text{Log}_e \left(\frac{q_i}{p_i} \right) \right)$$

$$I_{k,0}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \lim_{\alpha \rightarrow 0} \frac{1}{2\alpha + 1} \left(\sum_{i=1}^k \left(p_i \cdot \text{Log}_e \left(\frac{p_i}{q_i} \right) \cdot \left(\frac{p_i}{q_i} \right)^\alpha \right) \right) = \sum_{i=1}^k \left(p_i \cdot \text{Log}_e \left(\frac{p_i}{q_i} \right) \right)$$

3.4.1. Relações entre outras divergências direcionadas generalizadas

No caso da divergência direcionada generalizada proposta por Cresie e Read nenhum valor particular de α transforma esta na divergência direcionada de Kullback, porém no caso $\alpha = 0$, temos uma razão de diferença em termos da constante $\text{Log}_e 2$.

$$I_{k,0}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \frac{1}{\text{Log}_e 2} K(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k)$$

A divergência direcionada de ordem α proposta por Cressie e Read, pode ser vista em função das outras duas anteriormente citadas, através das seguintes equações:

$$I_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \frac{1}{\alpha \cdot (\alpha + 1)} \left(2^{\alpha \cdot I'_{k,\alpha+1}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k)} - 1 \right)$$

$$I_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k) = \frac{2^\alpha - 1}{\alpha \cdot (\alpha + 1)} I''_{k,\alpha+1}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k)$$

CAPÍTULO II

TESTES DE HIPÓTESES DE ADEQUABILIDADE DE AJUSTE EM DISTRIBUIÇÕES MULTINOMIAIS

1. Introdução

O modelo matemático do problema de comunicação, como visto em McMillan, B. (1953) [11], pode ser encarado de modo análogo com o problema de estimação. Uma mensagem é transmitida, porém no decorrer da transmissão temos interferências desconhecidas, alterando a mensagem recebida. Com base no que foi recebido temos que inferir ao que foi transmitido.

Podemos então fazer um paralelo com a inferência estatística, onde temos uma amostra, que admitimos ser proporcionada pela realização de alguma variável aleatória e queremos obter informações sobre o possível mecanismo aleatório gerador de tais dados, ou seja, desejamos conhecer tal variável aleatória.

Utilizando-se da Teoria de Informação, apresentada no capítulo precedente, vamos expandir a idéia de medida de divergência direcionada para uma aplicação em testes de hipóteses, como apresentado em Cressie, N. A. C. e Read, T. R. C. (1984) [5]. Tomando-se ao invés da medida de divergência direcionada entre duas distribuições de probabilidade de variáveis aleatórias discretas, uma medida de divergência entre o ponto observado, na realização de uma variável aleatória discreta com distribuição multinomial e o respectivo valor esperado desta.

2. Teste de hipóteses Multinomial

Seja Y uma variável aleatória discreta, ou discretizada, podendo assumir, com probabilidade não nula, os valores $\{ y_1, y_2, \dots, y_k \}$ com respectivas probabilidades dadas por: $P[Y = y_i] = \theta_i$, $i = 1, 2, \dots, k$. Definimos por espaço paramétrico, denotado por Δ , o conjunto de todos os possíveis valores dos θ_i 's

$$\Delta = \left\{ (\theta_1, \theta_2, \dots, \theta_k) : \theta_i > 0, i = 1, 2, \dots, k \text{ e } \sum_{i=1}^k \theta_i = 1 \right\}$$

Desejamos testar a hipótese de ser (p_1, p_2, \dots, p_k) ($(p_1, p_2, \dots, p_k) \in \Delta$) o possível valor do parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$. Para tanto tomamos uma amostra aleatória de tamanho n , dessa variável aleatória em questão, ou seja, geramos n valores independentes dessa variável aleatória. Seja $\{ Y_i \}$, $i = 1, 2, \dots, n$, essa amostra, temos que as variáveis aleatórias Y_i 's, são independentes e identicamente distribuídas (i.i.d.), onde a distribuição de cada Y_i é a mesma que Y .

Como visto em Bickel, P. J. e Doksum, K. A. (1977) [2], temos que a estatística (vetor) (X_1, X_2, \dots, X_k) , onde X_i corresponde ao número de vezes que ocorre o valor y_i na amostra acima mencionada, é uma estatística suficiente minimal em respeito ao parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$. Para testarmos algo em respeito ao parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$ será então suficiente analisarmos a estatística (vetor) (X_1, X_2, \dots, X_k) . Sabendo-se que a distribuição de (X_1, X_2, \dots, X_k) é multinomial de parâmetros $(n, (\theta_1, \theta_2, \dots, \theta_k))$, o problema de interesse, ou seja, testar se $(\theta_1, \theta_2, \dots, \theta_k) = (p_1, p_2, \dots, p_k)$ é reduzido a um teste multinomial.

Denotamos por Ω , o espaço amostral de (X_1, X_2, \dots, X_k) , temos:

$$\Omega = \left\{ (x_1, x_2, \dots, x_k) : x_i \in \{0, 1, \dots, n\}, i = 1, 2, \dots, k \text{ e } \sum_{i=1}^k x_i = n \right\}$$

O procedimento do teste será o de observar o valor produzido pela estatística (vetor) (X_1, X_2, \dots, X_k) , dado pelo ponto (x_1, x_2, \dots, x_k) $((x_1, x_2, \dots, x_k) \in \Omega)$ e posteriormente decidirmos se $(\theta_1, \theta_2, \dots, \theta_k)$ pode ser dado por (p_1, p_2, \dots, p_k) .

3. Breve histórico sobre testes de adequabilidade de ajuste em distribuições multinomiais

Um primeiro trabalho, que podemos citar como base para o desenvolvimento de testes de adequabilidade de ajuste em distribuições multinomiais, como citado em Lancaster, H. O. (1969) [9], é a normalidade assintótica, mencionada por Moivre e Laplace, da estatística:

$$\chi = (X - n.p) / \sqrt{n.p.(1-p)},$$

onde X é uma variável aleatória com distribuição Binomial, de parâmetros n (n inteiro positivo) e p ($p \in (0, 1)$), revelando a expressão:

$$\begin{aligned} \chi^2 &= (X - n.p)^2 / (n.p.(1-p)) = (X - n.p)^2 / (n.p) + (X - n.p)^2 / (n.(1-p)) = \\ &= (X - n.p)^2 / (n.p) + (n - X - n.(1-p))^2 / (n.(1-p)) \end{aligned}$$

que posteriormente foi generalizada por K. Pearson, para o caso de uma distribuição multinomial, ou seja:

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - n.p_i)^2}{n.p_i},$$

onde (X_1, X_2, \dots, X_k) é uma variável aleatória (vetor) com distribuição multinomial, de parâmetros n (n inteiro positivo) e (p_1, p_2, \dots, p_k) ($(p_1, p_2, \dots, p_k) \in \Delta$). Sugerindo que χ^2 tem distribuição assintótica Qui-quadrado com $k - 1$ graus de liberdade.

Uma grande contribuição para a aplicabilidade de χ^2 foi feita por R. A. Fisher, no caso onde existe uma necessidade de alguma estimação paramétrica. Para ser encontrado o valor de χ^2 , temos a distribuição assintótica de χ^2 Qui-quadrado com $k - s - 1$ graus de liberdade, onde s corresponde ao número de parâmetros estimados.

Em Cochran, W. G. (1952) [3], após uma minuciosa análise da estatística χ^2 apresentada anteriormente, temos a sugestão de um possível concorrente, a estatística do logaritmo da razão da verossimilhança, trazendo a luz outra estatística, com convergência Qui-quadrado, apresentando algumas propriedades, sob alguns aspectos, não encontradas em χ^2 .

Posteriormente muitos outros artigos estudaram essas estatísticas, onde em muitos desses trabalhos temos uma análise da semelhança existente entre essas duas estatísticas e, em alguns casos, encontramos novas sugestões de estatísticas alternativas a essas.

Alguns dos artigos que fazem comparação ou análises nesses testes com distribuição assintótica Qui-quadrado o fazem em amostras finitas, analisando casos particulares. Em muitos desses podemos notar uma dificuldade de aplicabilidade desses modelos de testes em pequenas amostras, principalmente em relação ao tamanho do teste que é sugerido e o real tamanho do teste.

Essas discussões prosseguem até os dias de hoje, com uma vasta literatura a esse respeito, e ainda irão prosseguir por muito tempo, pois não temos uma estatística, nesses moldes, que apresente um teste com uma forte característica, capaz de superar qualquer outra estatística. Sempre que analisamos o poder dos testes, podemos notar haver ora um, ora outro em melhor destaque, dependendo do ponto alternativo que estejamos analisando.

Em Cressie, N. A. C. e Read, T. R. C. (1984) [5], podemos ver grande parte dessas estatísticas, com convergência Qui-quadrado, generalizados em uma mesma família de estatísticas, incluindo-se a essas as estatísticas de K. Pearson e o logaritmo da razão de verossimilhança sugerido em Cochran, W. G. (1952) [3]. Porém, não contendo-se em generalizar essas estatísticas em uma família, Cressie e Read sugerem uma nova estatística, com convergência Qui-quadrado, dentro dessa família, como forte concorrente entre as demais já mencionadas,

$$2nl^{2/3} = \frac{9}{5} \sum_{i=1}^k \left\{ X_i \cdot \left[\left(\frac{X_i}{n \cdot p_i} \right)^{2/3} - 1 \right] \right\},$$

onde (X_1, X_2, \dots, X_k) é uma variável aleatória (vetor) com distribuição multinomial, de parâmetros n (n inteiro positivo) e (p_1, p_2, \dots, p_k) ($(p_1, p_2, \dots, p_k) \in \Delta$).

4. Testes multinomiais apresentados em Cressie, N. A. C. e Read, T. R. C. (1984) [5]

4.1. Introdução

Partindo-se da divergência direcionada de ordem α de Cressie e Read, apresentada no capítulo anterior, $I_{k,\alpha}(p_1, p_2, \dots, p_k; q_1, q_2, \dots, q_k)$, pode-se encarar um teste de hipóteses multinomial como uma aplicação dessa divergência, onde o ponto (q_1, q_2, \dots, q_k) representa o valor do parâmetro que queremos testar e (p_1, p_2, \dots, p_k) uma estimativa, baseada na observação de (x_1, x_2, \dots, x_k) , para o valor do parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$.

Uma outra maneira de utilização da idéia de medida de divergência direcionada, como é sugerida em Cressie, N. A. C. e Read, T. R. C. (1984) [5], é observarmos o valor da divergência entre o valor esperado da variável aleatória multinomial em questão, supondo que o valor do parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$ é dado por (q_1, q_2, \dots, q_k) e o valor observado na realização dessa variável (x_1, x_2, \dots, x_k) . Temos nessa medida um caráter direcional, pois estamos analisando a divergência entre um ponto esperado, de uma suposta variável com distribuição multinomial e um ponto observado dentro do espaço amostral relativo a esta variável.

4.2. Função de divergência apresentada em Cressie, N. A. C. e Read, T. R. C. (1984) [5]

Em Cressie, N. A. C. e Read, T. R. C. (1984) [5] e posteriormente em Read, T. R. C. e Cressie, N. A. C. (1988) [14] nos é apresentado uma família de estatísticas construídas por intermédio de uma variável aleatória (vetor) com distribuição multinomial de parâmetros n e $(\theta_1, \theta_2, \dots, \theta_k)$, onde n é um número inteiro não negativo e $(\theta_1, \theta_2, \dots, \theta_k) \in \Delta$. Essa família é definida como segue:

$$2nI^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^\lambda - 1 \right]; \lambda \in \mathfrak{R} \text{ (números reais)}$$

onde: (X_1, X_2, \dots, X_k) é uma variável aleatória (vetor) com distribuição multinomial, como acima citado, (E_1, E_2, \dots, E_k) é o valor (vetor) esperado dessa variável aleatória e λ o parâmetro indexador dessa família de estatísticas.

Para os valores $\lambda = -1$ e $\lambda = 0$, $2nI^\lambda$ é definido por continuidade, ou seja, pelos respectivos limites. Temos então:

$$\begin{aligned} 2nI^{-1} &= \lim_{\lambda \rightarrow -1} 2nI^\lambda = \lim_{\lambda \rightarrow -1} \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^\lambda - 1 \right] = \lim_{\lambda \rightarrow -1} \frac{2}{\lambda^2 + \lambda} \sum_{i=1}^k X_i \left[e^{\lambda \log_e \left(\frac{X_i}{E_i} \right)} - 1 \right] = \\ &= \lim_{\lambda \rightarrow -1} \frac{2}{2\lambda + 1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left[e^{\lambda \log_e \left(\frac{X_i}{E_i} \right)} - 1 \right] = \lim_{\lambda \rightarrow -1} \frac{2}{2\lambda + 1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left(\frac{X_i}{E_i} \right)^\lambda = \\ &= \frac{2}{-2+1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left(\frac{X_i}{E_i} \right)^{-1} = -2 \sum_{i=1}^k E_i \log_e \left(\frac{X_i}{E_i} \right) = 2 \sum_{i=1}^k E_i \log_e \left(\frac{E_i}{X_i} \right). \end{aligned}$$

$$\begin{aligned}
2nl^0 &= \lim_{\lambda \rightarrow 0} 2nl^\lambda = \lim_{\lambda \rightarrow 0} \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^\lambda - 1 \right] = \lim_{\lambda \rightarrow 0} \frac{2}{\lambda^2 + \lambda} \sum_{i=1}^k X_i \left[e^{\lambda \log_e \left(\frac{X_i}{E_i} \right)} - 1 \right] = \\
&= \lim_{\lambda \rightarrow 0} \frac{2}{2\lambda + 1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left[e^{\lambda \log_e \left(\frac{X_i}{E_i} \right)} - 1 \right] = \lim_{\lambda \rightarrow 0} \frac{2}{2\lambda + 1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left(\frac{X_i}{E_i} \right)^\lambda = \\
&= \frac{2}{0+1} \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right) \left(\frac{X_i}{E_i} \right)^0 = 2 \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right).
\end{aligned}$$

Essa família de estatísticas, para um número k fixo, possui distribuição assintótica Qui-quadrado com $k-1$ graus de liberdade, para qualquer valor de λ . Essa afirmação é provada em dois passos, primeiramente, utilizando uma expansão em série de Taylor, pode ser provado que a distribuição da estatística $2nI^\lambda$ é semelhante a distribuição dessa estatística no ponto $\lambda = 1$, ou seja, $2nI^1$. O segundo passo é tão somente a identificação de $2nI^1$ como a estatística de Pearson, que é já bastante conhecida como distribuída assintoticamente Qui-quadrado, com $k-1$ graus de liberdade.

4.2.1. Casos particulares de $2nI^\lambda$

Para cada valor de λ , na família de estatísticas apresentada anteriormente, temos uma estatística diferente. Observando-se alguns valores de λ podemos notar, nessa família de estatísticas, além da estatística de Pearson, já mencionada, algumas estatísticas bastante conhecidas, dentre elas podemos destacar as que seguem:

4.2.1.1. Estatística X^2 modificada de Neymann ($\lambda = -2$)

$$\begin{aligned}
2nl^{-2} &= \frac{2}{-2(-2+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^{-2} - 1 \right] = \sum_{i=1}^k \left(\frac{E_i^2}{X_i} - X_i \right) = \sum_{i=1}^k \left(\frac{E_i^2 - X_i^2}{X_i} \right) \\
&= \sum_{i=1}^k \left(\frac{E_i^2 - X_i^2 + 2X_i E_i - 2X_i E_i + X_i^2 - X_i^2}{X_i} \right) = \sum_{i=1}^k \left(\frac{E_i^2 - 2X_i E_i + X_i^2 + 2X_i E_i - 2X_i^2}{X_i} \right) = \\
&= \sum_{i=1}^k \frac{(X_i - E_i)^2 + 2(X_i E_i - X_i^2)}{X_i} = \sum_{i=1}^k \frac{(X_i - E_i)^2}{X_i} + 2 \sum_{i=1}^k (E_i - X_i) = \sum_{i=1}^k \frac{(X_i - E_i)^2}{X_i}
\end{aligned}$$

4.2.1.2. Estatística do logaritmo da razão de verossimilhança modificado ($\lambda = -1$)

$$2nl^{-1} = \lim_{\lambda \rightarrow -1} 2nl^\lambda = 2 \sum_{i=1}^k E_i \log_e \left(\frac{E_i}{X_i} \right).$$

4.2.1.3. Estatística de Freeman - Tukey ($\lambda = -0.5$)

$$\begin{aligned}
2nl^{-0.5} &= \frac{2}{-0.5(-0.5+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^{-0.5} - 1 \right] = -8 \sum_{i=1}^k (\sqrt{X_i E_i} - X_i) = 4 \sum_{i=1}^k (-2\sqrt{X_i E_i} + 2X_i) = \\
&= 4 \sum_{i=1}^k (-2\sqrt{X_i E_i} + E_i + X_i) = 4 \sum_{i=1}^k (\sqrt{X_i} - \sqrt{E_i})^2
\end{aligned}$$

4.2.1.4. Estatística do logaritmo da razão de verossimilhança ($\lambda = 0$)

$$2nl^0 = \lim_{\lambda \rightarrow 0} 2nl^\lambda = 2 \sum_{i=1}^k X_i \log_e \left(\frac{X_i}{E_i} \right).$$

4.2.1.5. Estatística X^2 de Pearson ($\lambda = 1$)

$$\begin{aligned} 2nl^1 &= \frac{2}{1(1+1)} \sum_{i=1}^k X_i \left[\left(\frac{X_i}{E_i} \right)^1 - 1 \right] = \sum_{i=1}^k \left(\frac{X_i^2}{E_i} - X_i \right) = \sum_{i=1}^k \left(\frac{X_i^2 - X_i E_i}{E_i} \right) = \\ &= \sum_{i=1}^k \left(\frac{X_i^2 - X_i E_i - X_i E_i + X_i E_i - E_i^2 + E_i^2}{E_i} \right) = \sum_{i=1}^k \left(\frac{X_i^2 - 2X_i E_i + E_i^2 + X_i E_i - E_i^2}{E_i} \right) = \\ &= \sum_{i=1}^k \frac{(X_i - E_i)^2}{E_i} + \sum_{i=1}^k (X_i - E_i) = \sum_{i=1}^k \frac{(X_i - E_i)^2}{E_i} \end{aligned}$$

4.2.2. Utilização da família de estatísticas em testes multinomiais

Em cada valor de λ temos uma função que de alguma maneira tenta quantificar uma divergência entre o valor esperado (vetor) (E_1, E_2, \dots, E_k) e o ponto (vetor), do espaço amostral Ω , observado (x_1, x_2, \dots, x_k) na realização da variável aleatória com distribuição multinomial, associada ao experimento em questão.

Temos, sob algum ponto de vista matemático, que é refletido em diferentes valores de λ , uma divisão, dentro do espaço amostral Ω , entre pontos com maior plausibilidade à hipótese nula, que levariam a sua não rejeição, caso fossem observados nessa realização, acima descrita e pontos com

menor plausibilidade à hipótese nula, que levariam a sua rejeição, caso fossem observados nessa realização. Muitos desses diferentes valores de λ , citado anteriormente, foram tratados separadamente, criando-se um certo número de testes de adequabilidade de ajuste com distribuição assintótica Qui-quadrado. Na família de estatísticas proposta por Cressie e Read, temos uma generalização de muitos desses, podendo analisá-los como casos particulares.

Observamos então, para cada valor de λ , uma função que indica a divergência entre o ponto observado, na realização da variável aleatória multinomial e o respectivo valor (vetor) esperado dessa variável.

4.2.3. Comparação assintótica entre os testes da família Cressie e Read

Temos em 4.2. que as estatísticas de divergência $2nI^\lambda$, ($\lambda \in \mathfrak{R}$) são assintoticamente equivalentes, sob a hipótese nula considerada como verdadeira, com distribuição assintótica comum, Qui-quadrado com $k-1$ graus de liberdade. A eficiência assintótica dos testes, utilizando essas estatísticas, para um número de células k fixo, utilizando o conceito de eficiência relativa de Pitman e Bahadur, para algumas certas alternativas, foi mostrado em Cressie, N. A. C. e Read, T. R. C. (1984) [5], não haver grandes diferenças, ou seja, as estatísticas da família $2nI^\lambda$, mostraram-se, nesses casos analisados, assintoticamente equivalentes em relação à eficiência em alguns casos particulares de alternativas.

Em relação à velocidade de convergência das estatísticas de divergência $2nI^\lambda$, ($\lambda \in \mathfrak{R}$) o método de comparação utilizado em Cressie, N. A. C. e Read, T. R. C. (1984) [5] foi o de observar, na expansão assintótica dos momentos exatos da família $2nI^\lambda$, utilizando série de Taylor, o valor de segunda ordem, fazendo-se modificações nas variáveis, de sorte a obtermos no valor de primeira ordem o respectivo momento de uma variável aleatória com distribuição Qui-quadrado com $k-1$ graus de liberdade. Observando-se para grandes valores de n , o valor de segunda ordem, em relação a essas expansões, para os três primeiros momentos, mostra-se mais próximo a zero, nos casos de $\lambda = 1$ e $\lambda = 2/3$, sugerindo uma boa velocidade de convergência nesses dois casos, bem como mostra-se muito próximo a zero, para $\lambda \in [0.3 , 2.7]$.

Observamos então o surgimento de mais um valor de λ ($\lambda = 2/3$), para ser estudado entre as estatísticas, generalizadas em $2nI^\lambda$, ($\lambda \in \mathfrak{R}$) para sabermos qual é a melhor em testes de adequabilidade de ajustes em distribuições multinomiais.

4.2.4. Comparação em amostras finitas entre os testes da família Cressie e Read

O critério utilizado em Cressie, N. A. C. e Read, T. R. C. (1984) [5], para comparação entre os testes utilizando as estatísticas em $2nI^\lambda$, ($\lambda \in \mathfrak{R}$) é o critério da função poder, onde em muitos casos, devido à complexidade de hipóteses alternativas, é matematicamente inacessível. Para essa comparação foram utilizados casos particulares onde a hipótese nula é a simétrica ($1/k, \dots, 1/k$), com $k = 4$, $n = 20$ e tamanho do teste aleatorizado 0.05, tomando-se alternativas fazendo-se uma perturbação em uma das coordenadas, tomando-se as coordenadas restantes com valores iguais.

Os resultados mostram-se de tal modo que para alternativas onde a perturbação é feita para valores maiores que $1/k$, o poder do teste cresce com o valor de λ e de modo contrário, para alternativas onde a perturbação é feita para valores menores que $1/k$, o poder do teste decresce com o valor de λ .

Podemos notar que é possível, para alternativas particulares, melhorarmos o poder de testes já bastante discutidos, como o teste definido por Pearson bem como o teste da razão do logaritmo da verossimilhança, escolhendo outros membros da família $2nI^\lambda$.

CAPÍTULO III

SIMULAÇÃO PARA ALGUNS CASOS PARTICULARES

1. Análise em pequenas amostras

Como podemos observar em Tate, M. W. e Hyer, L. A. (1973) [18] a imprecisão do teste Qui-quadrado fica bastante evidente em amostras de pequeno porte. Dependendo da relação de n , tamanho da amostra e k , número de classes ou categorias, podemos ter uma grande diferença entre o tamanho do teste que é sugerido e o tamanho real do teste. Isso fica mais acentuado ainda, quando o valor associado à hipótese nula p , está longe do caso equiprovável ($1/k, 1/k, \dots, 1/k$).

Uma das idéias encontradas na ampla literatura, como alternativa para a aplicação do teste Qui-quadrado é defendida no artigo Tate, M. W. e Hyer, L. A. (1973) [18]. A sugestão encontrada nesse trabalho é a de que não devemos empregar o teste Qui-quadrado nos casos de amostras de pequeno porte, tendo como alternativa, usar um teste de tamanho "exato" partindo-se das probabilidades de ser verificado cada ponto do espaço amostral, considerando que a hipótese nula seja verdadeira. A região de rejeição, desse teste, é construída tomando-se os pontos de menores probabilidades, sob a hipótese nula considerada como verdadeira, até a obtenção do tamanho de teste desejado. Temos como procedimento do teste a rejeição da hipótese nula, sob um nível de significância α , caso o somatório das probabilidades de valores, que apresentam probabilidades menores ou iguais ao ponto observado, seja menor do que o valor α .

Uma outra idéia, não tão radical, em relação ao abandono do teste Qui-quadrado, quanto à primeira, é defendida no artigo Radlow, R e Alf Jr. , E. F. (1975) [12]. A sugestão encontrada nesse trabalho é que devemos utilizar a estatística χ^2 , considerando a hipótese nula como verdadeira, apenas para ordenar os valores dentro do espaço amostral. A região de rejeição desse teste é

constituída pelos pontos que apresentam os maiores valores para a estatística χ^2 , considerando a hipótese nula como verdadeira, até a obtenção do tamanho de teste desejado. Temos como procedimento do teste a rejeição da hipótese nula, sob um nível de significância α , caso o somatório das probabilidades de valores, que apresentam o valor da estatística χ^2 maior ou igual ao valor da estatística χ^2 aplicada no ponto observado, seja menor do que o valor de α .

Em ambas alternativas ainda devemos ter, na maioria dos casos, uma diferença entre o tamanho do teste que é sugerido e o real tamanho do teste. Essa diferença só será completamente extinta, caso façamos uma aleatorização no teste, ou seja, incluirmos na região de rejeição alguns pontos, a partir de cada critério, acima mencionado, com uma certa probabilidade de rejeição da hipótese nula, caso seja verificado esse ponto, adequada essa probabilidade de tal modo que o tamanho real do teste seja o valor sugerido, α .

2. Aplicação da idéia de Radlow, R e Alf Jr. , E. F. (1975) [12] em testes da família definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5].

De modo análogo ao teste Qui-quadrado, os testes da família definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5], possuem uma grande imprecisão para pequenas amostras. Dependendo da relação de n , tamanho da amostra e k , número de classes ou categorias, podemos ter uma grande diferença entre o tamanho do teste que é sugerido, com o tamanho real do teste, isso fica mais acentuado ainda, quando o valor associado à hipótese nula p , está longe do caso equiprovável ($1/k$, $1/k$, ..., $1/k$) e/ou quando o valor de λ , na função de divergência, é ou muito pequeno ($\lambda \rightarrow -\infty$) ou muito grande ($\lambda \rightarrow +\infty$).

Usando a sugestão encontrada em Radlow, R e Alf Jr. , E. F. (1975) [12], porém aleatorizando o teste, podemos estabelecer mecanismos de comparação entre os testes dessa família, com mesmo tamanho, em se tratando de pequenas amostras, inferindo em possíveis resultados para grandes amostras.

3. Análise de alguns casos particulares em testes definidos através da função de divergência definido em Cressie, N. A. C. e Read, T. R. C. (1984) [5].

A família de estatísticas definidas em Cressie, N. A. C. e Read, T. R. C. (1984) [5], gera uma infinidade de testes, além disso temos infinita, não enumerável, maneiras de optarmos pela hipótese nula e infinitas maneiras de selecionamos n , tamanho da amostra e k , número de classes ou categorias. Tomamos, para uma análise, uma simulação de alguns casos particulares, porém com uma certa representatividade dos demais casos.

i) Optamos pelo modelo trinomial ($k = 3$), pois poderíamos ver em gráficos, com grande facilidade, as regiões de rejeição.

ii) Tomamos tamanhos de amostras tais que não fossem muito pequenos, para obtermos uma boa visualização da região de rejeição, porém não muito grandes para não extrapolar o tempo computacional de execução do programa criado para a construção das regiões de rejeição ($n \in \{ 20, 30, 35 \}$).

iii) Optamos por hipóteses nulas que pudessem ser um pouco representativas de todo o espaço amostral ($(\theta_1, \theta_2, \theta_3) \in \{ (0.1, 0.2, 0.7), (0.2, 0.3, 0.5), (1/3, 1/3, 1/3) \}$).

iv) Tomamos o tamanho do teste de 0.05 porque, além de ser um tamanho tradicionalmente usado em toda a literatura de testes de adequabilidade de ajustes, é um valor que apresenta uma região de rejeição contendo um razoável número de pontos, para os tamanhos de amostra utilizados e possui uma margem de significância aceitável, em testes de hipóteses.

v) Tomamos os valores de λ em $\{ -5, -1, 0, 2/3, 1, 5 \}$ pois: -5 representa um valor de tal modo que as regiões de rejeição com valores de λ menores, não apresentam muita diferença deste valor; -1

representa o teste do logaritmo da razão de verossimilhança modificado; 0 representa o teste do logaritmo da razão de verossimilhança; $2/3$ o teste com valor sugerido por Cressie e Read; 1 representa o teste Qui-quadrado, com sugerido por Pearson; 5 representa um valor de tal modo que as regiões de rejeição com valores de λ maiores, não apresentam muita diferença deste valor; além disso essa lista de valores de λ cobre uma certa região contínua entre -5 e 5, com pequenas diferenças entre as regiões de rejeição apresentadas por eles.

4. Procedimentos para a construção das regiões de rejeição

4.1. Teste Qui-quadrado, aplicado na função de divergência definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5]

Denotamos por Teste Qui-quadrado, aplicado na função de divergência definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5] nesse caso, os testes construídos por intermédio de um percentil da distribuição Qui-quadrado. Como, para qualquer valor de λ na função de divergência, definida por Cressie e Read, temos uma convergência para a distribuição Qui-quadrado, construímos uma região formada por pontos que fazem parte da região de rejeição da hipótese nula como em um teste Qui-quadrado comum, ou seja, caso o valor da função de divergência, aplicado no ponto em questão, seja superior a um certo percentil da distribuição Qui-quadrado com $k-1$ graus de liberdade, temos esse ponto pertencendo a essa região de rejeição. Temos então essas regiões formadas por pontos que apresentam valores, da função de divergência, acima mencionada, sob a hipótese nula em questão, para os determinados valores de λ , maiores que o valor crítico $\chi^2 = 5.9915$, onde esse valor é tal que: $P[\chi_2^2 > 5.9915] \cong 0.05$, onde χ_2^2 representa a distribuição Qui-quadrado com dois graus de liberdade.

4.2. Teste "Qui-quadrado" como sugerido em Radlow, R e Alf Jr. , E. F. (1975) [12], usando a função de divergência definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5]

Essas regiões são formadas por pontos que apresentam os maiores valores da função de divergência, para o respectivo valor de λ em questão, de tal modo que a soma das probabilidades desses pontos, sob a respectiva hipótese nula em questão, não ultrapassasse o valor do tamanho do teste (0.05), ou seja, tomamos os pontos pertencentes ao espaço amostral que apresentam as maiores divergências em relação ao valor esperado, sob a respectiva hipótese nula em questão, de modo que a soma de suas probabilidades não ultrapassem ao valor 0.05

4.3. Aleatorização

Como estamos tratando de uma variável aleatória discreta, para alcançarmos o real tamanho do teste como é proposto necessitamos de uma aleatorização. Os pontos aleatorizados são pontos destinados a gerar o tamanho real do teste de 0.05, a respectiva probabilidade associada a esses pontos é a probabilidade de rejeição da hipótese nula em questão, caso seja verificado esse ponto, na respectiva realização da variável aleatória em questão.

Obs: Podemos ter regiões formadas por pontos que pertencem simultaneamente às regiões definidas em 4.1. e 4.2. (pág. 39 e 40), bem como podemos ter pontos aleatorizados (4.3. pág. 40) que pertencem à região definida em 4.1. (pág. 39).

5. Máscara de leitura dos gráficos relativos às simulações

Os gráficos que serão vistos nas páginas seguintes foram construídos para mostrar as regiões de rejeição apresentadas nos testes escolhidos para a simulação dos casos particulares mencionados: distribuição trinomial ($k = 3$), com hipóteses nulas dadas por $\{ (0.1, 0.2, 0.7), (0.2, 0.3, 0.5), (1/3, 1/3, 1/3) \}$, para os tamanhos de amostras $n \in \{ 20, 30, 35 \}$, para os valores de λ , da função de divergência de Cressie e Read, dados por $\{ -5, -1, 0, 2/3, 1, 5 \}$, para o tamanho de teste 0.05

i) Variável aleatória (vetor) em questão: $(X_1, X_2, X_3) \sim \text{Multinomial} (n, (\theta_1, \theta_2, \theta_3))$

ii) Hipóteses em questão

$H_0 : (\theta_1, \theta_2, \theta_3) = (q_1, q_2, q_3)$ $H_A : (\theta_1, \theta_2, \theta_3) \neq (q_1, q_2, q_3)$

iii) Valor de λ : valor empregado na função de divergência de Cressie e Read

iv) Ponto aleatorizado: Os pontos aleatorizados são pontos destinados a gerar o tamanho real do teste de 0.05

v) Gráfico:

Título I (i.): Hipótese nula em questão

Título II (i.j.): Tamanho do teste

Título III (i.j.h.): Valor de λ , da função de divergência de Cressie e Read, e os pontos aleatorizados com respectivas probabilidades de rejeição à hipótese nula em questão.

Eixos: x, valor da coordenada X_1 e y, valor da coordenada X_2 correspondentes ao espaço amostral da variável aleatória com distribuição multinomial em questão.

Pontos:

() região definida em 4.1. pág. 39 () região definida em 4.2. pág. 40

() pontos definidos em 4.3. pág. 40

() pontos que pertencem simultaneamente as regiões definidas em 4.1. pág. 39 e 4.2. pág. 40

() pontos que pertencem simultaneamente as regiões definidas em 4.1. pág. 39 e 4.3. pág. 40

i. Relaciona a hipótese nula

i.j. Relaciona o tamanho da amostra

i.j.h. Relaciona o valor de λ , da função de divergência de Cressie e Read, e os pontos aleatorizados com sua respectiva probabilidade de rejeição à hipótese nula em questão

Eixo y, valor da coordenada X2 da variável multinomial em questão

○ pontos definidos em 4.2. pág. 40

X pontos definidos em 4.3. pág. 40

✕ pontos definidos simultaneamente em 4.2. e 4.3. pág. 40

● pontos definidos simultaneamente em 4.1. pág. 39 e 4.2. pág. 40

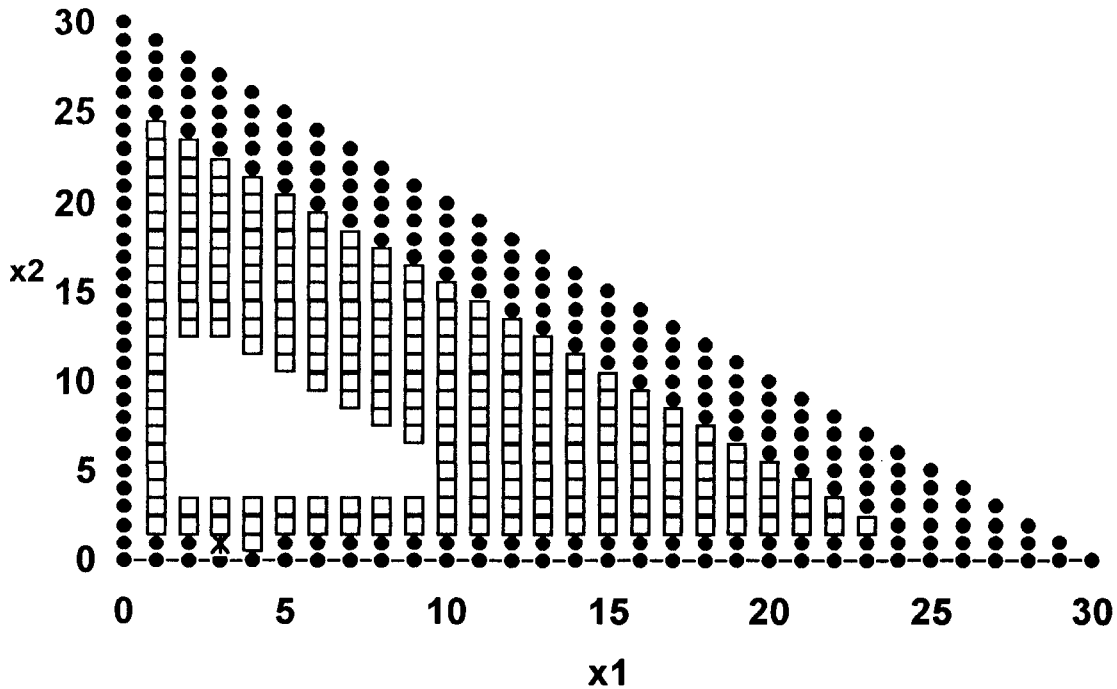
□ pontos definidos em 4.1. pág. 39

Eixo x, valor da coordenada X1 da variável multinomial em questão

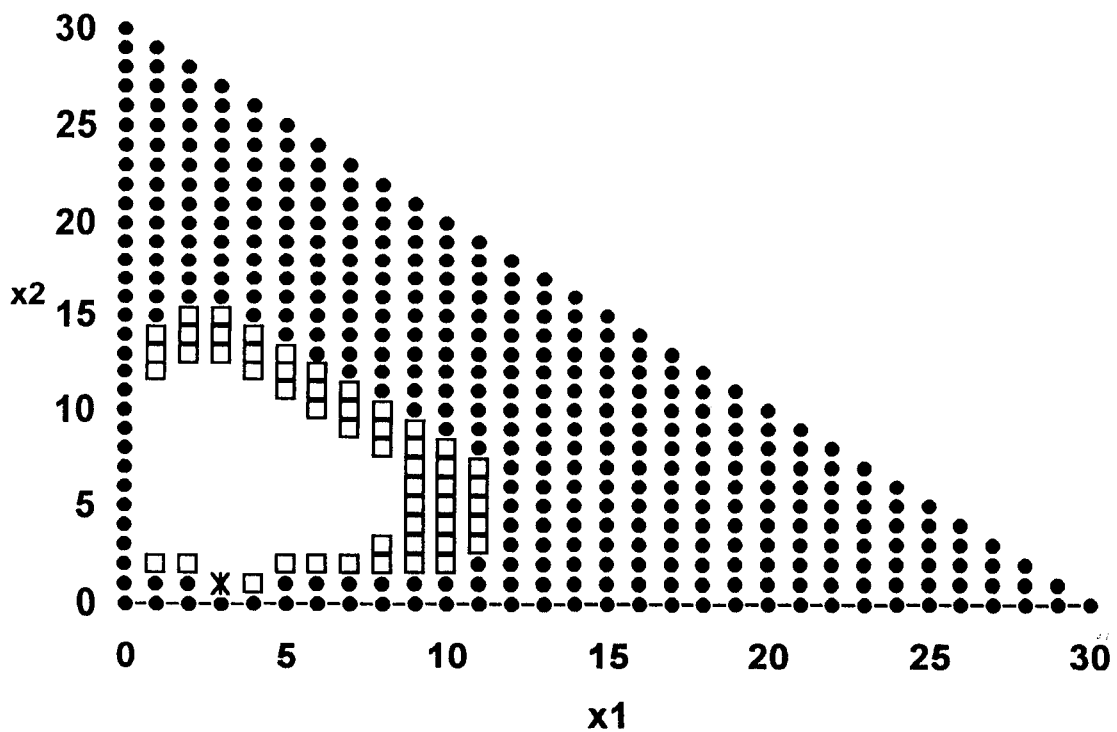
6. Hipótese nula dada por $(0.1, 0.2, 0.7)$.

6.1. Tamanho da amostra 30.

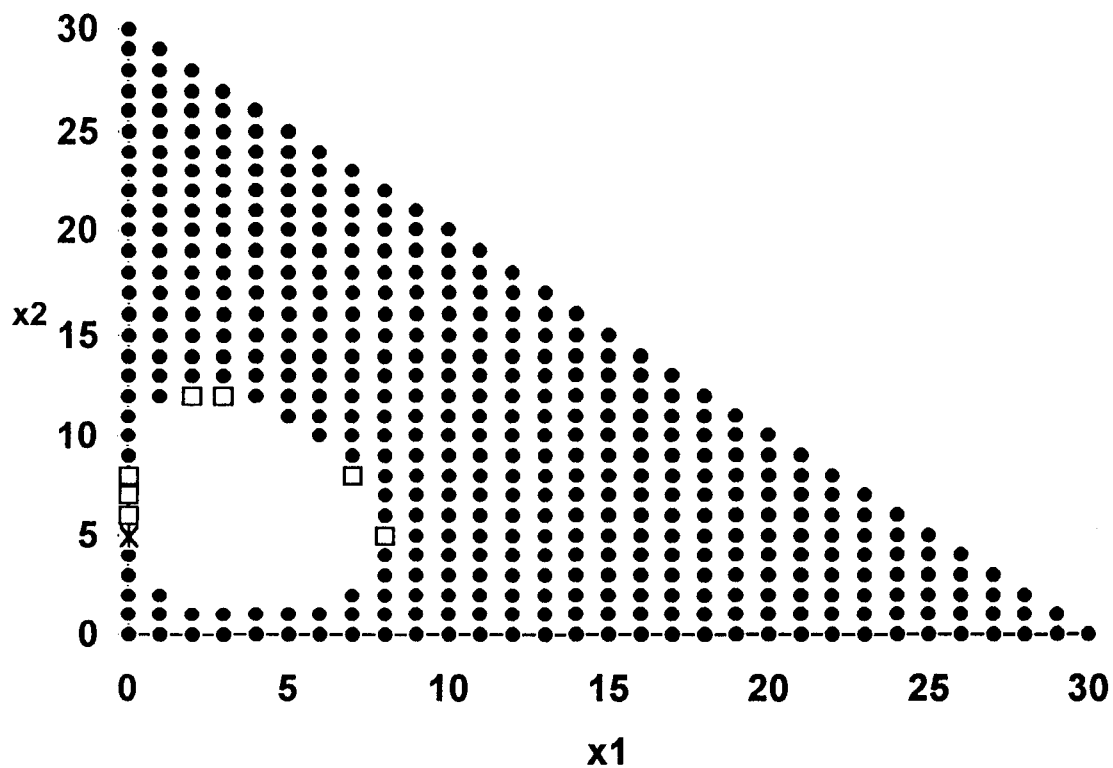
6.1.1. $\lambda = -5$, ponto aleatorizado $(3, 1, 26)$, com probabilidade aproximada 0.617702



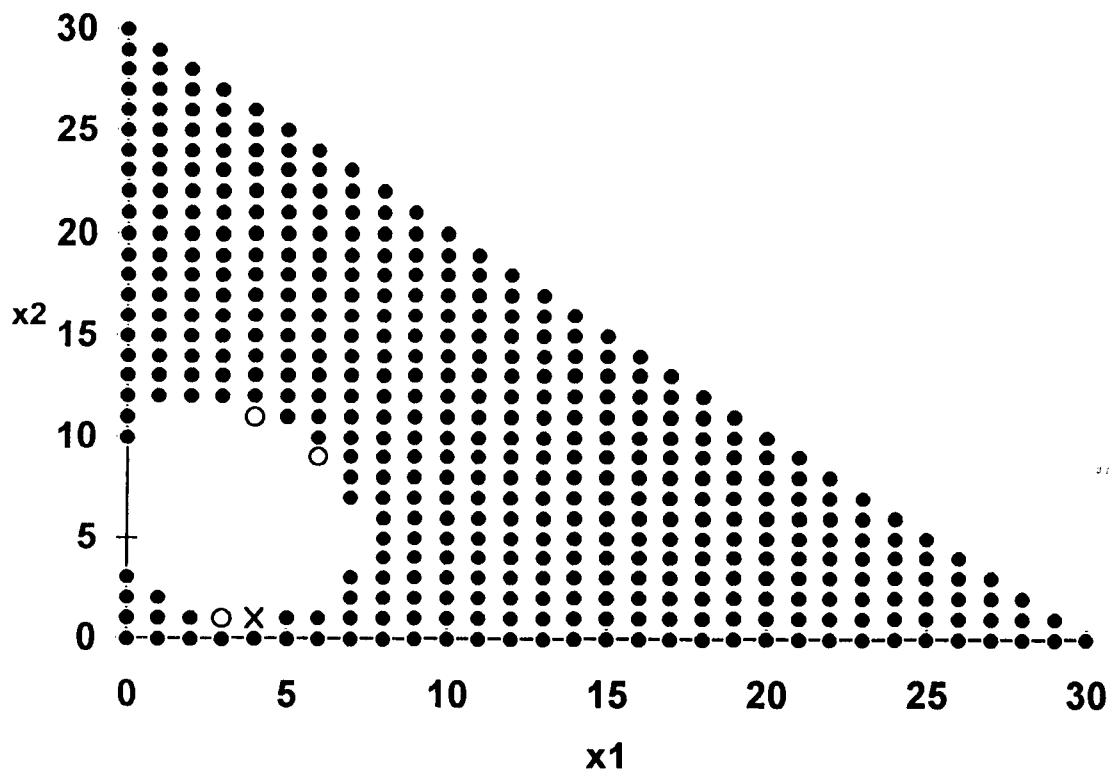
6.1.2. $\lambda = -1$, ponto aleatorizado $(3, 1, 26)$, com probabilidade aproximada 0.480912



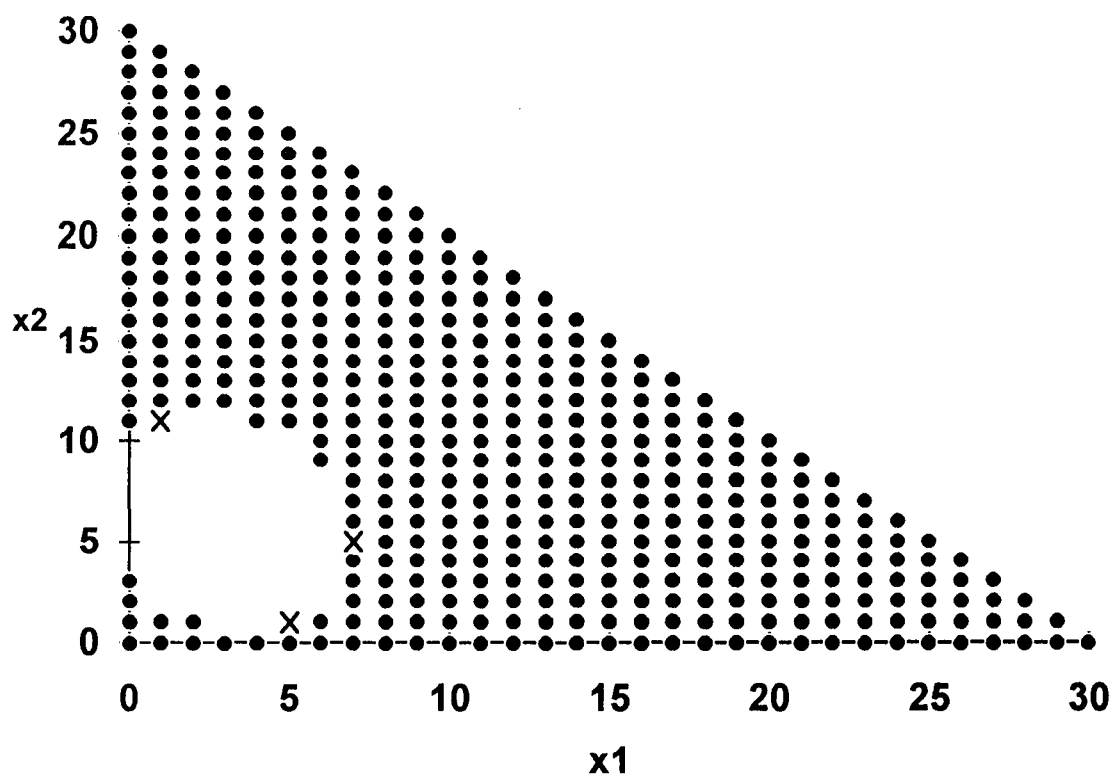
6.1.3. $\lambda = 0$, ponto aleatorizado (0 , 5 , 25), com probabilidade aproximada 0.764531



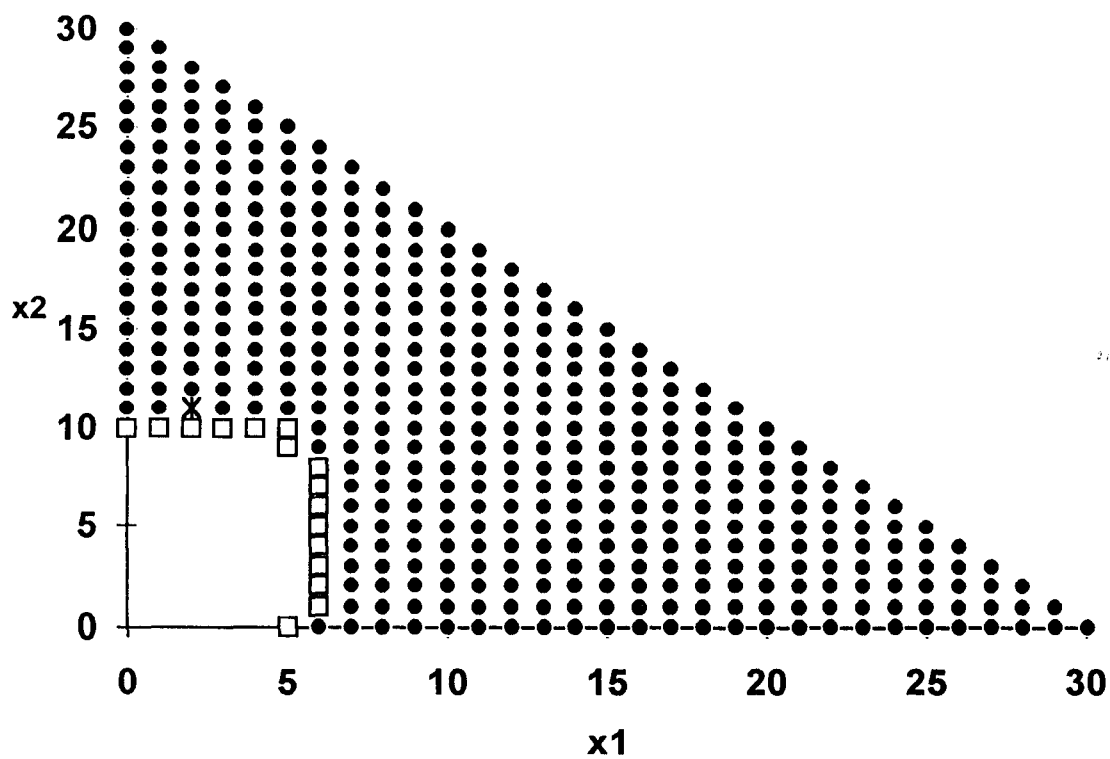
6.1.4. $\lambda = 2/3$, ponto aleatorizado (4 , 1 , 25), com probabilidade aproximada 0.465489



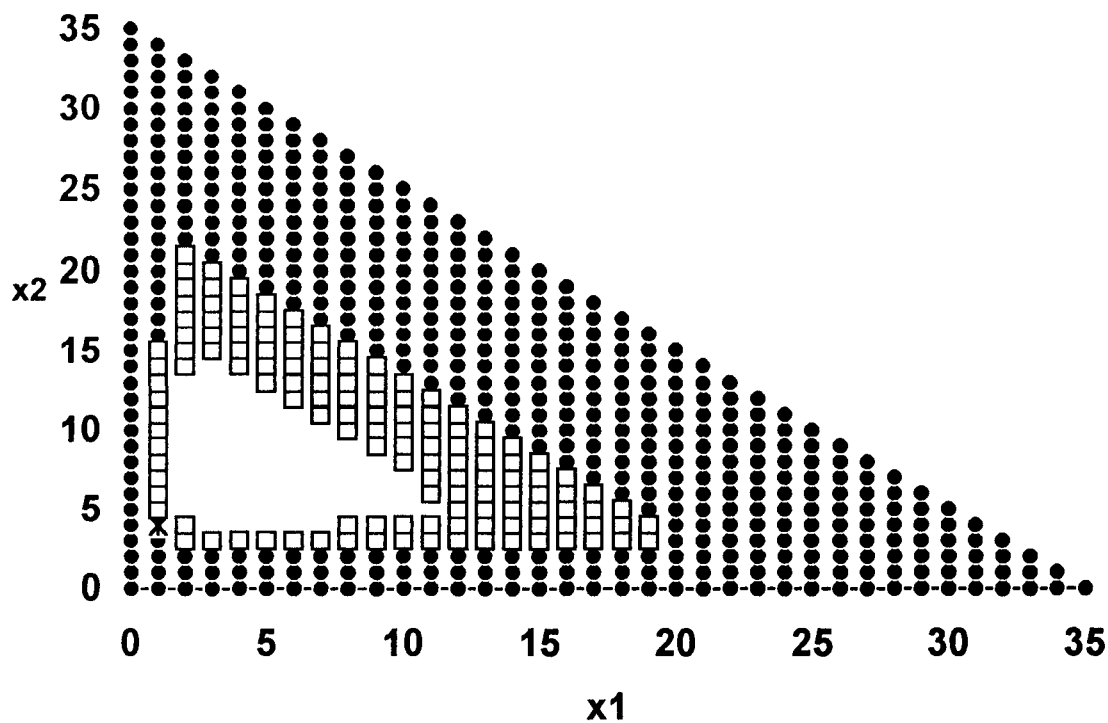
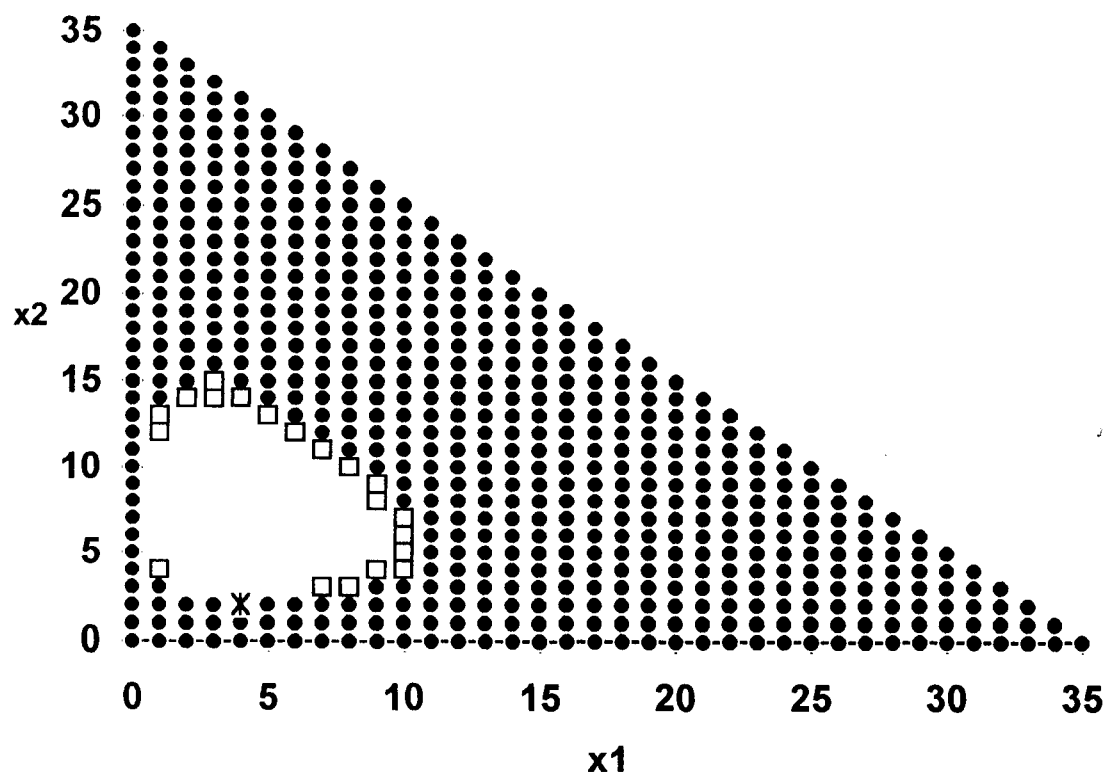
6.1.5. $\lambda = 1$, pontos aleatorizados $\{ (5 , 1 , 24) , (1 , 11 , 18) , (7 , 5 , 18) \}$, com probabilidade aproximada 0.431353



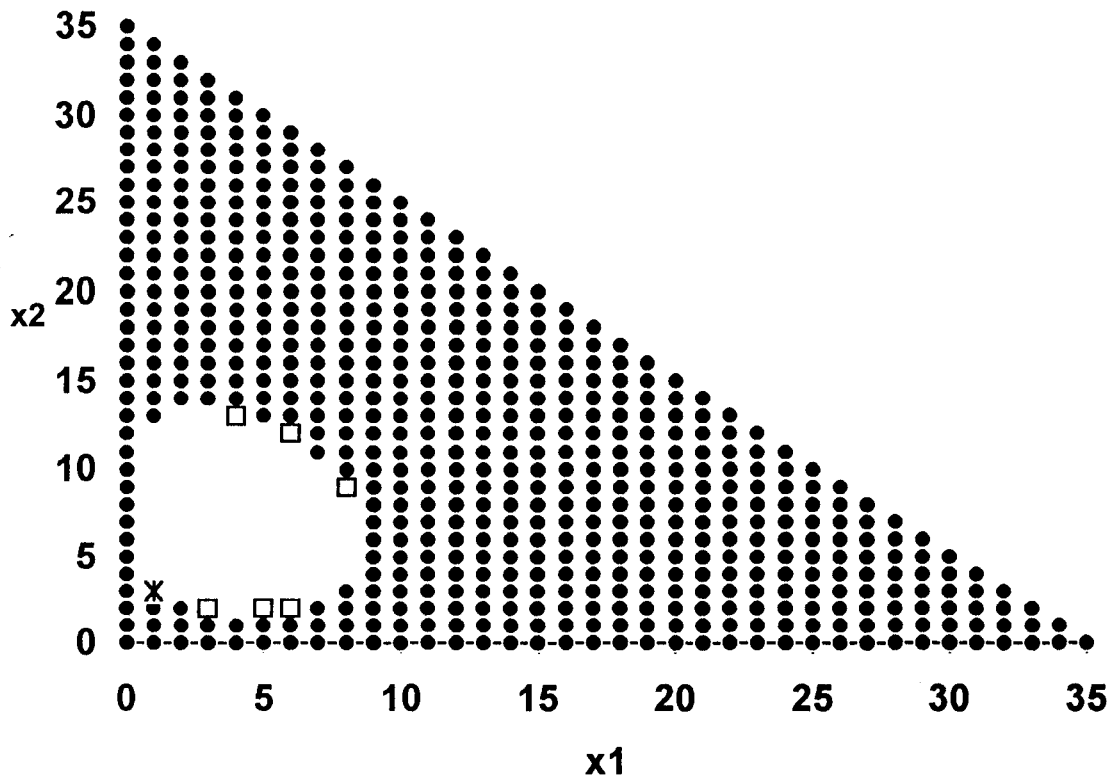
6.1.6. $\lambda = 5$, ponto aleatorizado $(2 , 11 , 17)$, com probabilidade aproximada 0.0744596



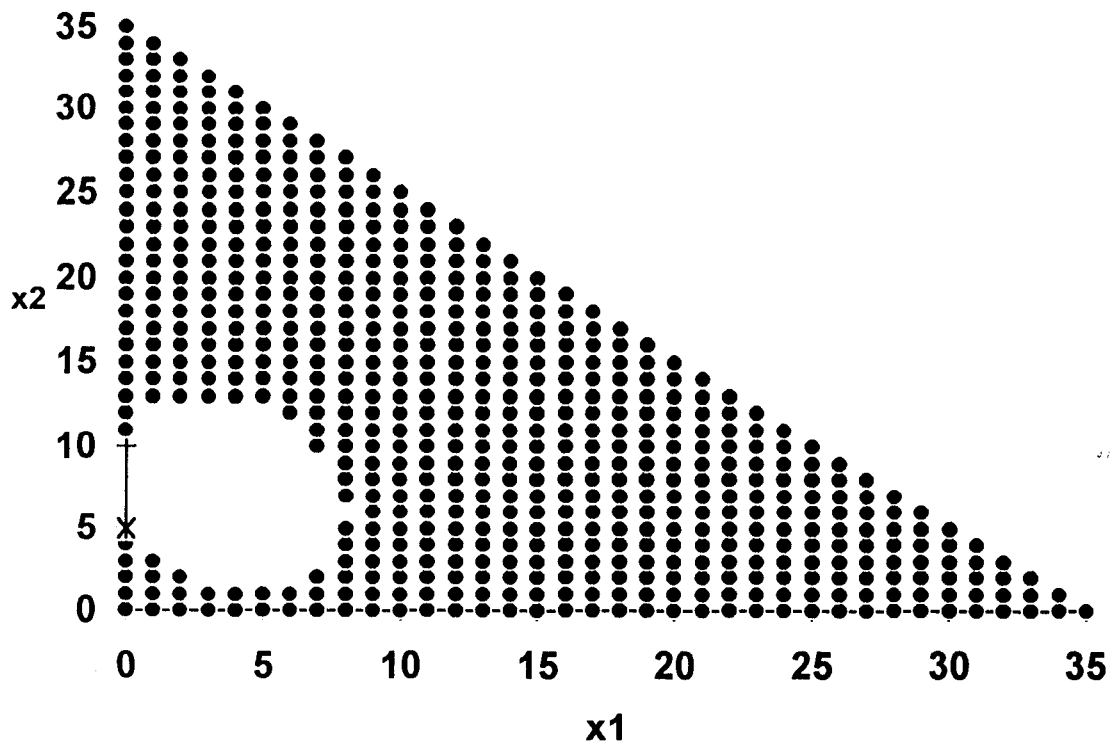
6.2. Tamanho da amostra 35.

6.2.1. $\lambda = -5$, ponto aleatorizado (1 , 4 , 30), com probabilidade aproximada 0.5794056.2.2. $\lambda = -1$, ponto aleatorizado (4 , 2 , 29), com probabilidade aproximada 0.613748

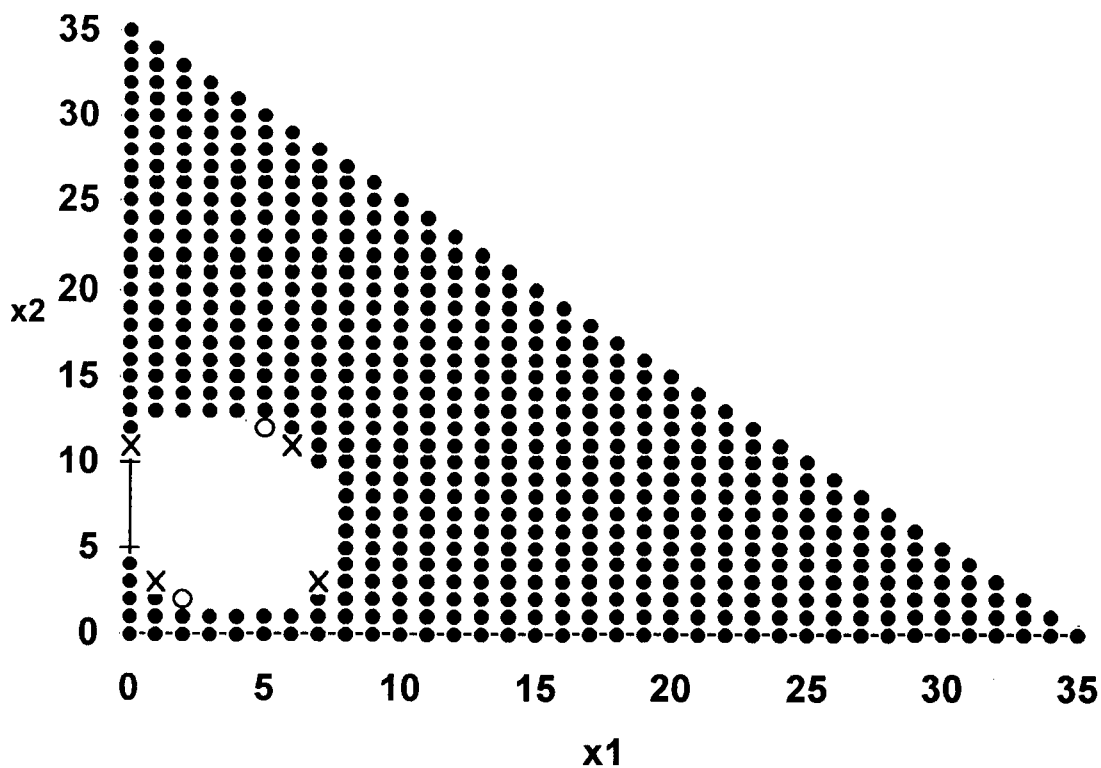
6.2.3. $\lambda = 0$, ponto aleatorizado (1 , 3 , 31), com probabilidade aproximada 0.0904701



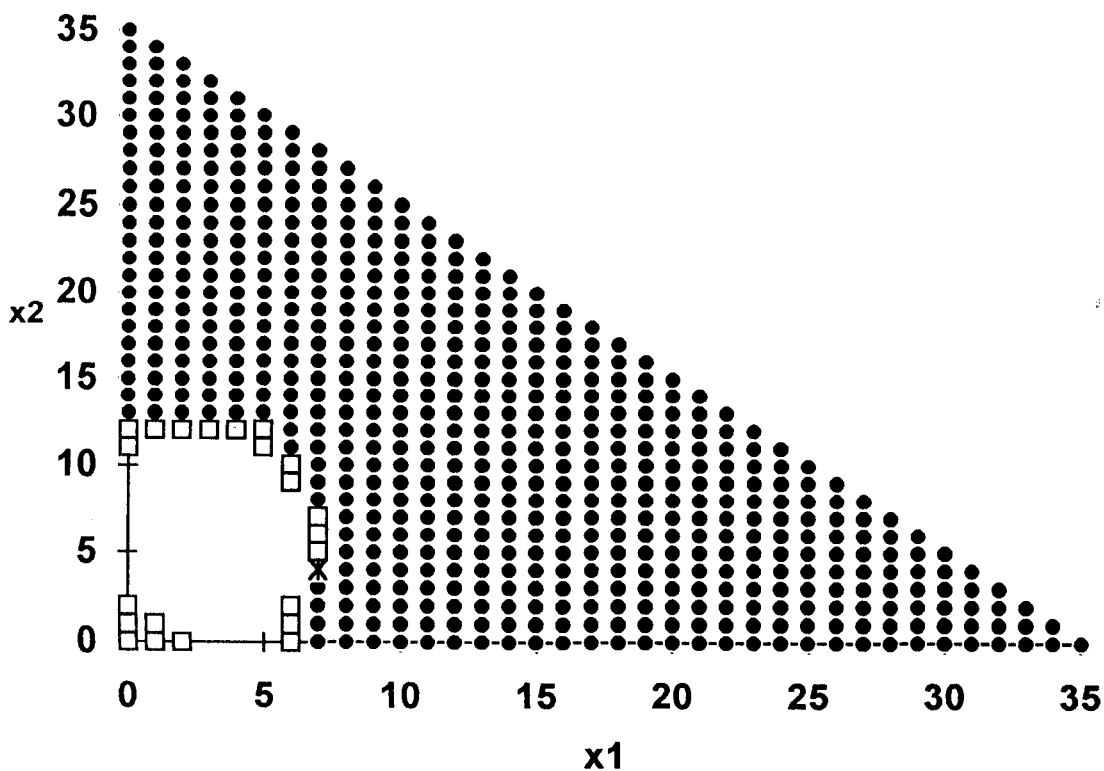
6.2.4. $\lambda = 2/3$, ponto aleatorizado (0 , 5 , 30), com probabilidade aproximada 0.120898



6.2.5. $\lambda = 1$, pontos aleatorizados $\{ (7, 3, 25); (0, 11, 24), (1, 3, 31), (6, 11, 18) \}$, com probabilidade aproximada 0.0235291



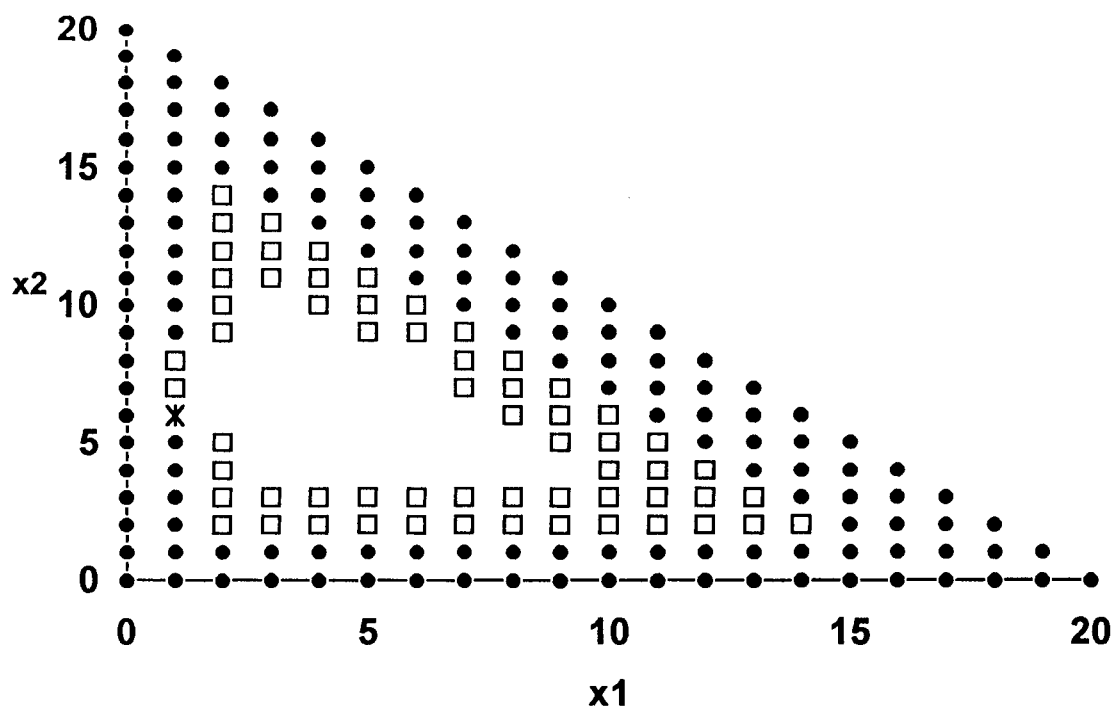
6.2.6. $\lambda = 5$, ponto aleatorizado $(7, 4, 24)$, com probabilidade aproximada 0.487348



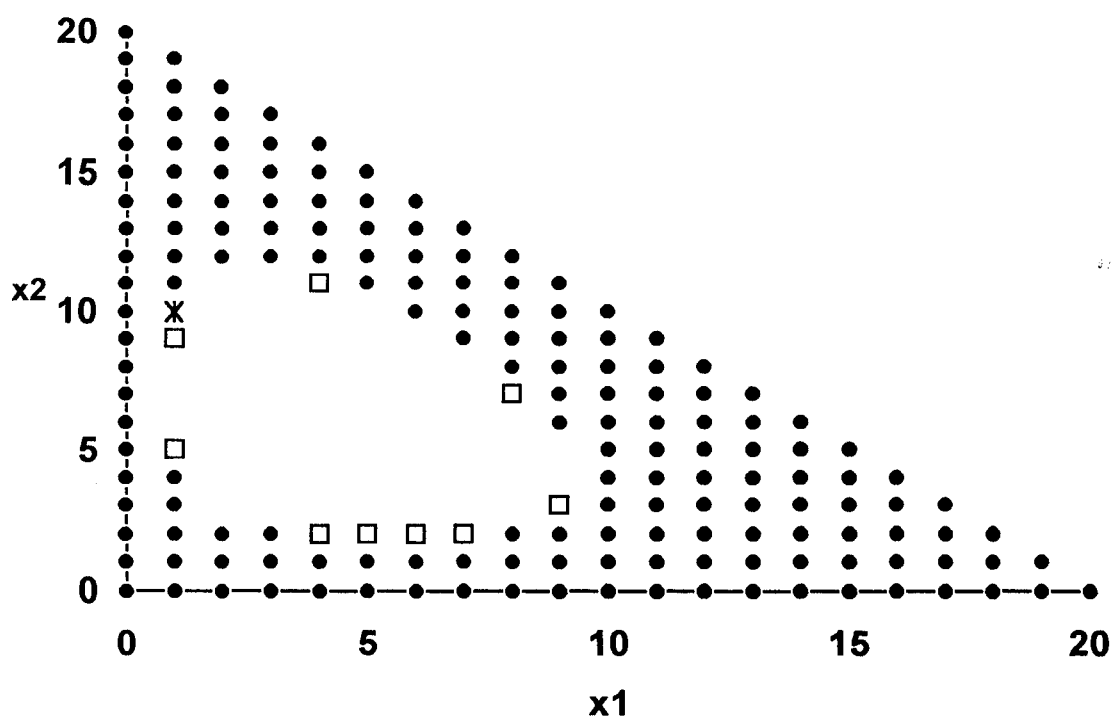
7. Hipótese nula dada por $(0.2, 0.3, 0.5)$.

7.1. Tamanho da amostra 20.

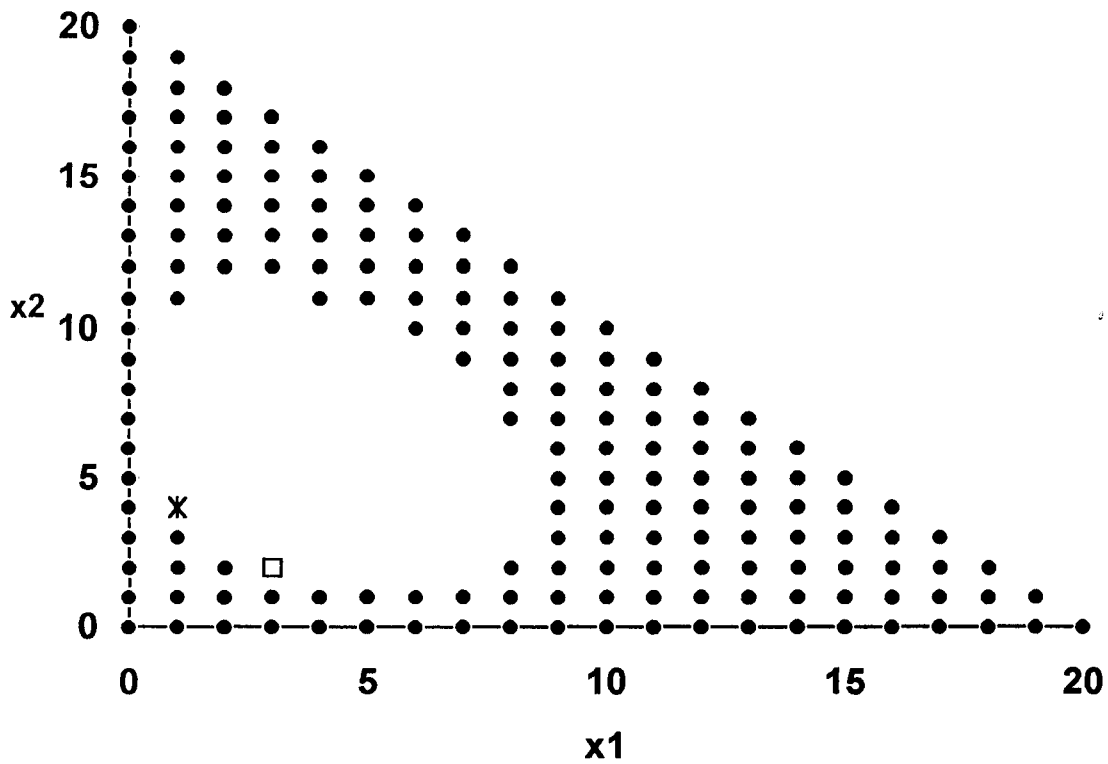
7.1.1. $\lambda = -5$, ponto aleatorizado $(1, 6, 13)$, com probabilidade aproximada 0.218827



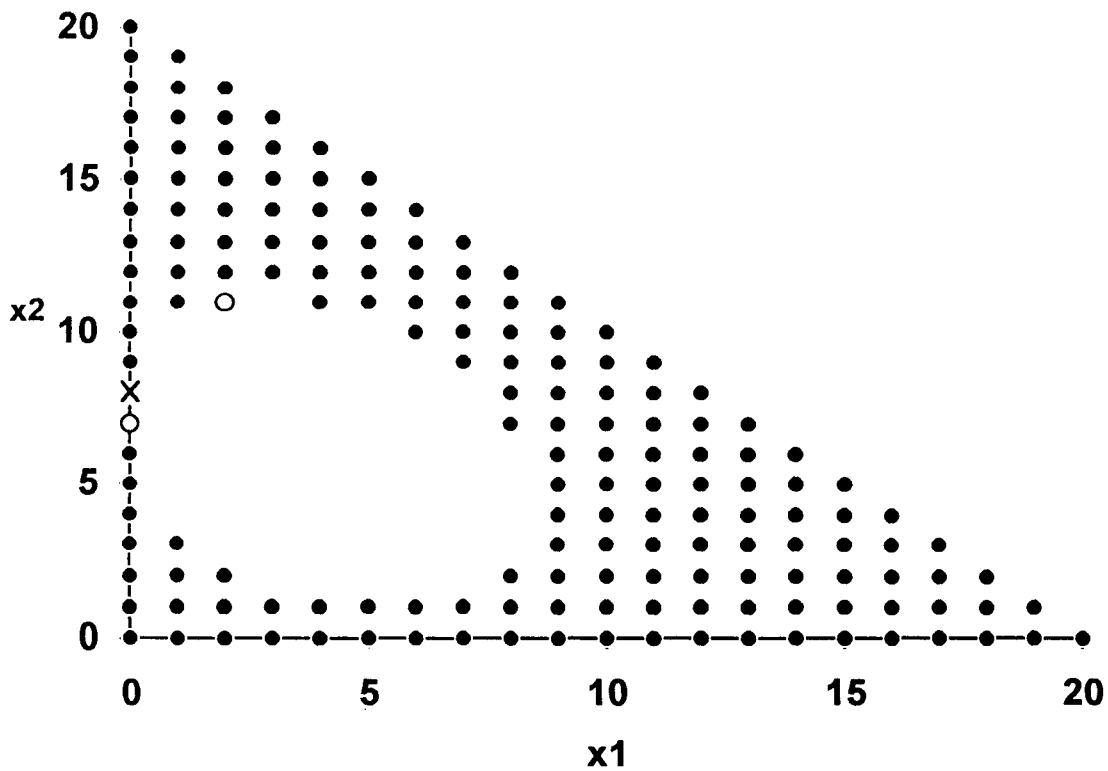
7.1.2. $\lambda = -1$, ponto aleatorizado $(1, 10, 9)$, com probabilidade aproximada 0.637488



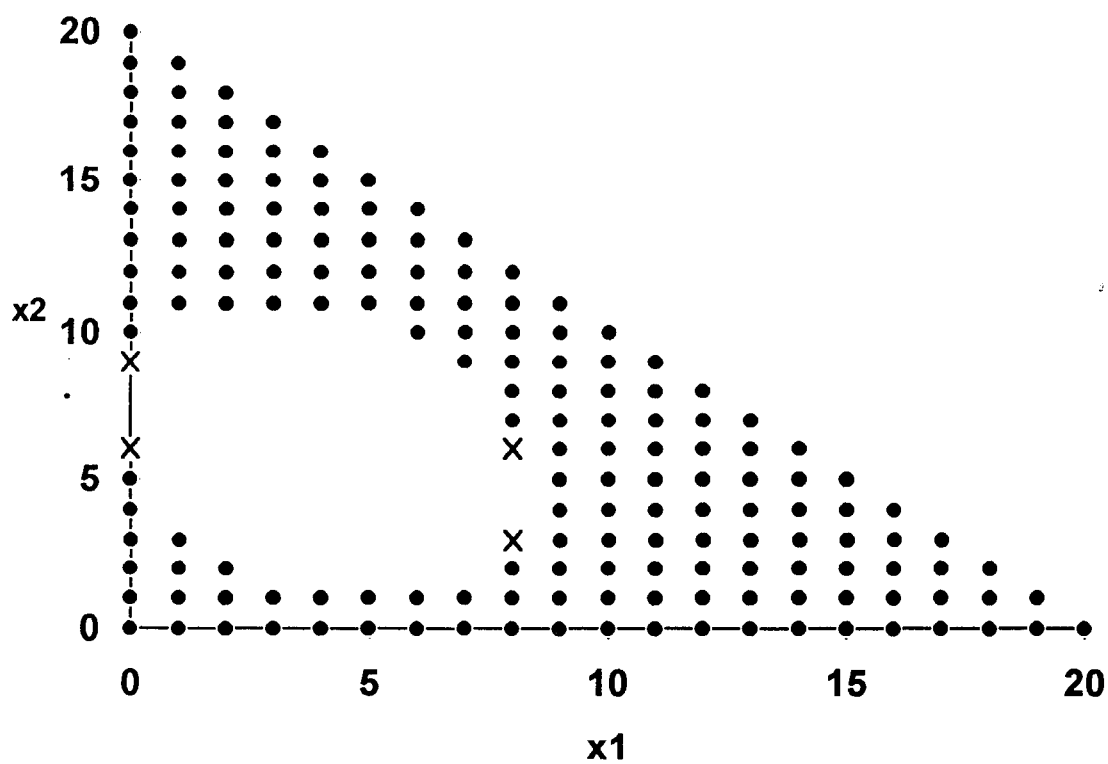
7.1.3. $\lambda = 0$, ponto aleatorizado (1, 4, 15), com probabilidade aproximada 0.400512



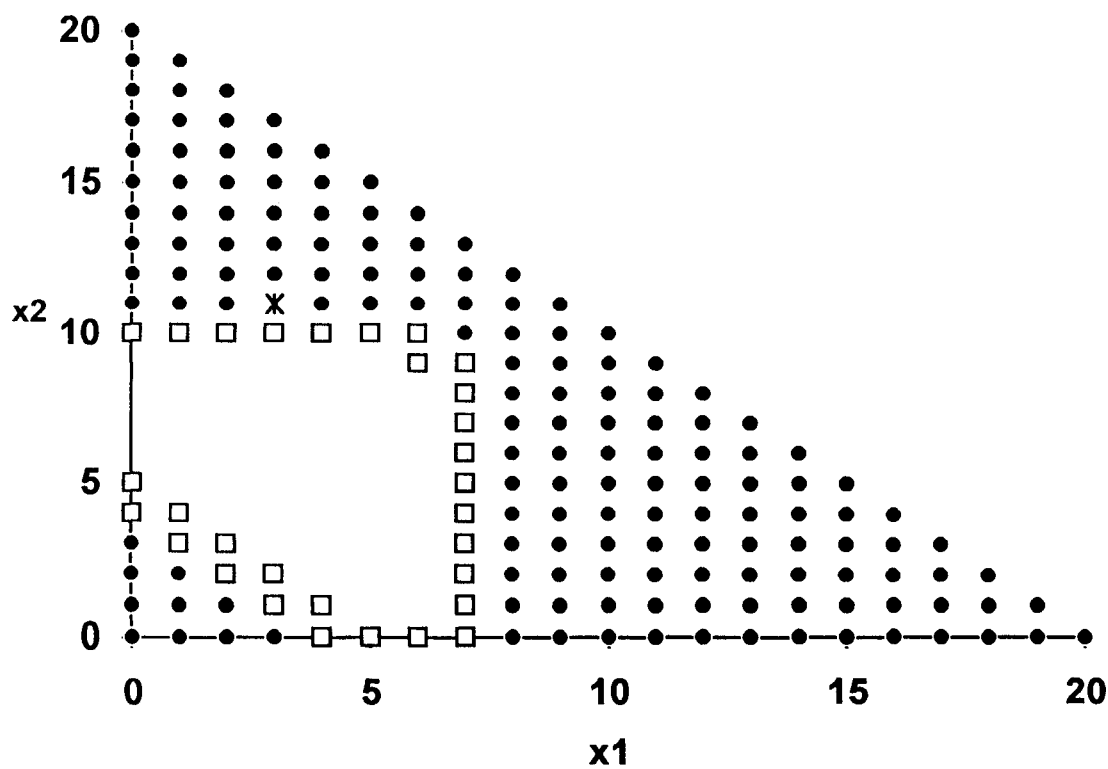
7.1.4. $\lambda = 2/3$, ponto aleatorizado (0, 8, 12), com probabilidade aproximada 0.101827



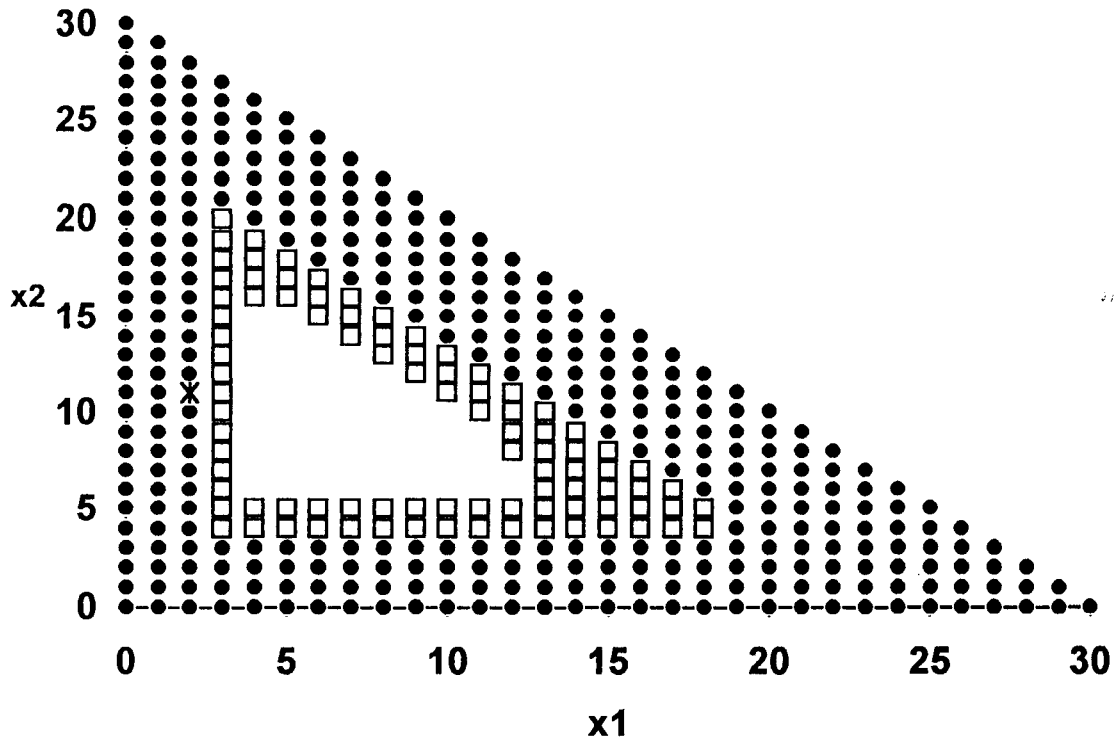
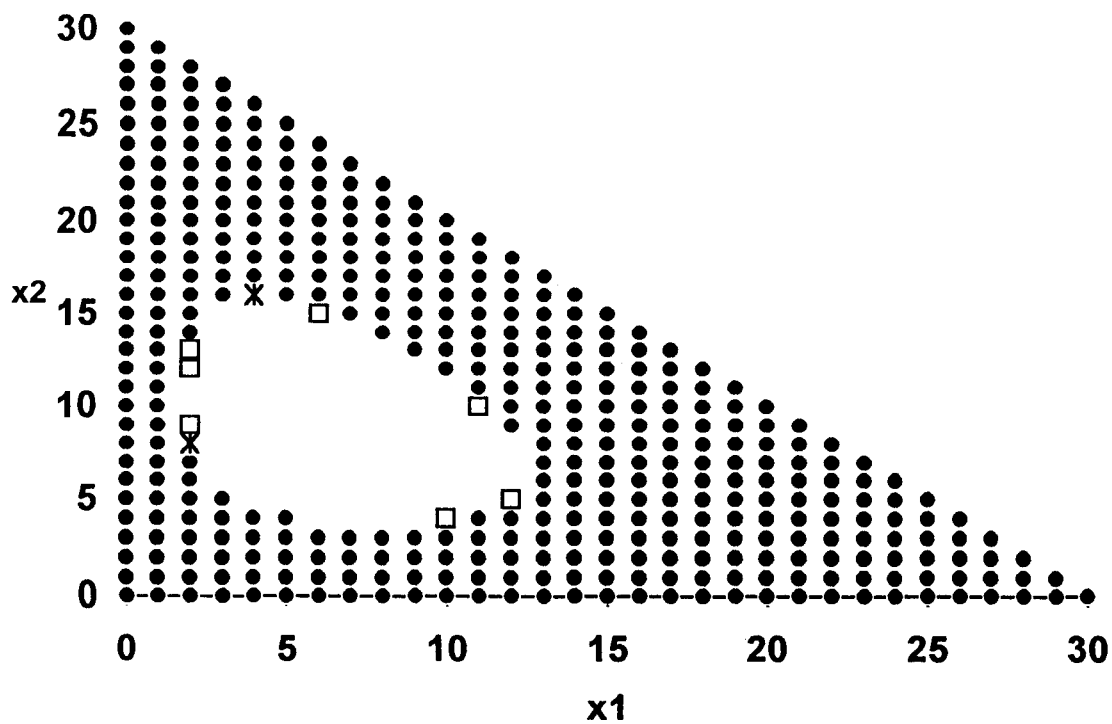
7.1.5. $\lambda = 1$, pontos aleatorizados $\{(0, 6, 14), (0, 9, 11), (8, 3, 9), (8, 6, 6)\}$, com probabilidade aproximada 0.237699



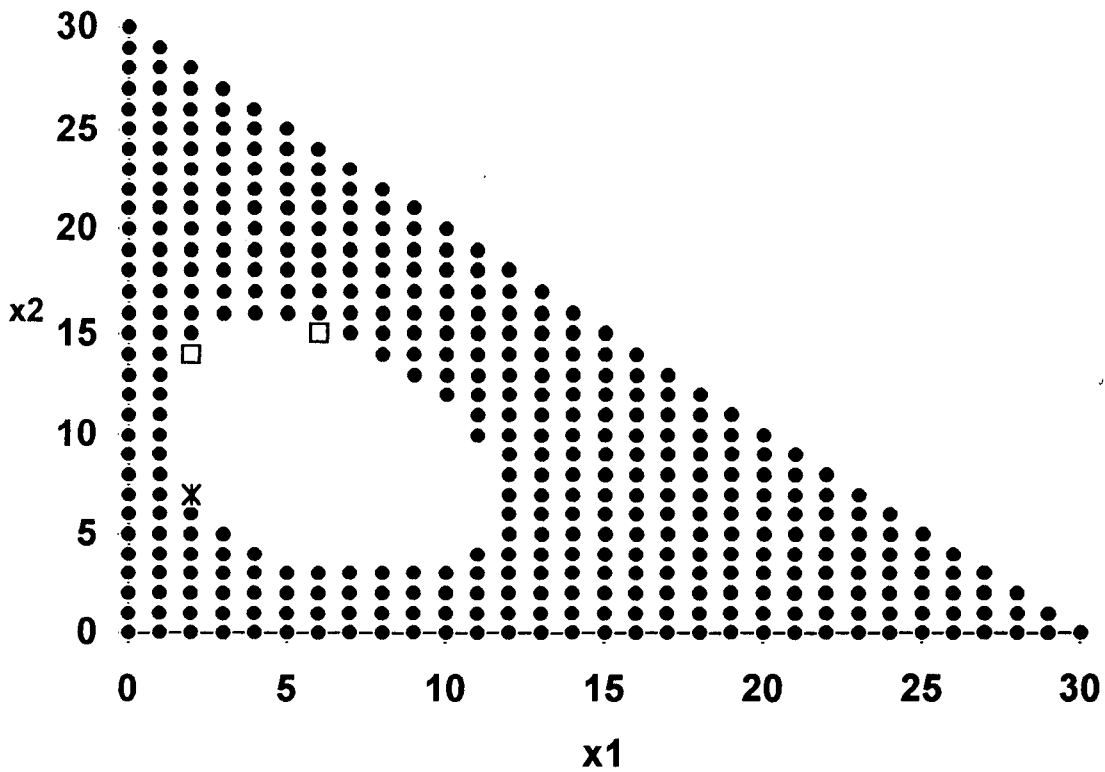
7.1.6. $\lambda = 5$, ponto aleatorizado $(3, 11, 6)$, com probabilidade aproximada 0.749816



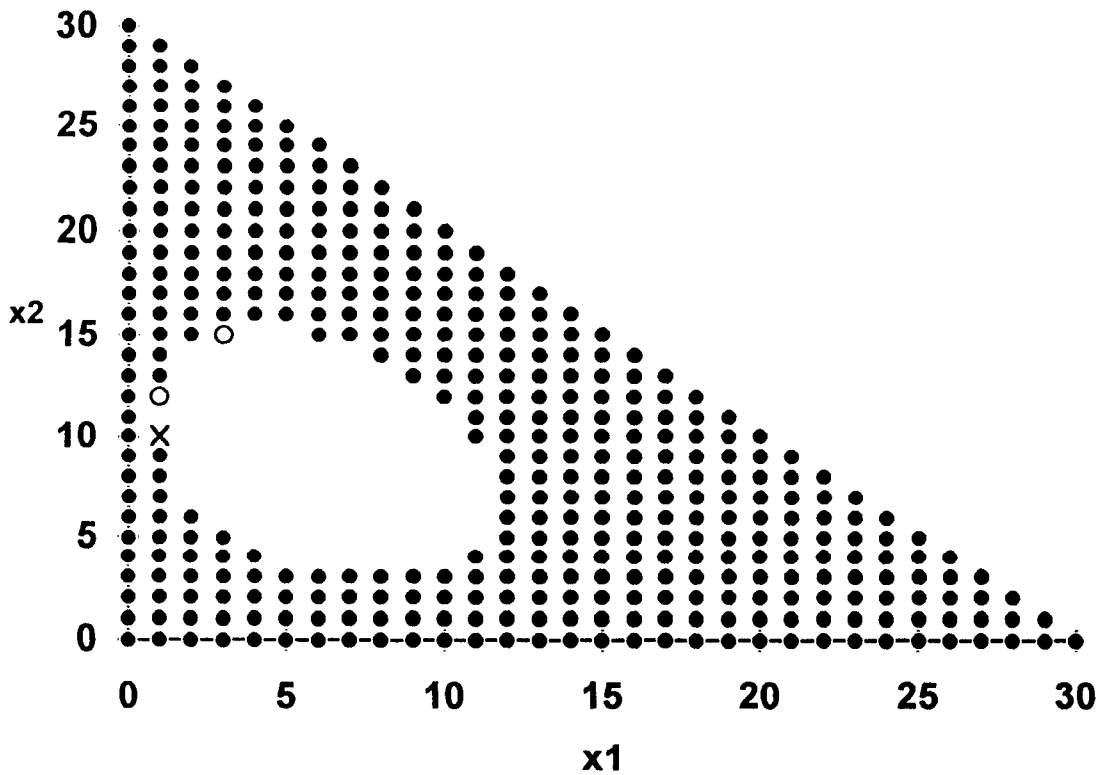
7.2. Tamanho da amostra 30.

7.2.1. $\lambda = -5$, ponto aleatorizado $(2, 11, 17)$, com probabilidade aproximada 0.1795717.2.2. $\lambda = -1$, pontos aleatorizados $\{ (2, 8, 20), (4, 16, 10) \}$, com probabilidade aproximada 0.330634

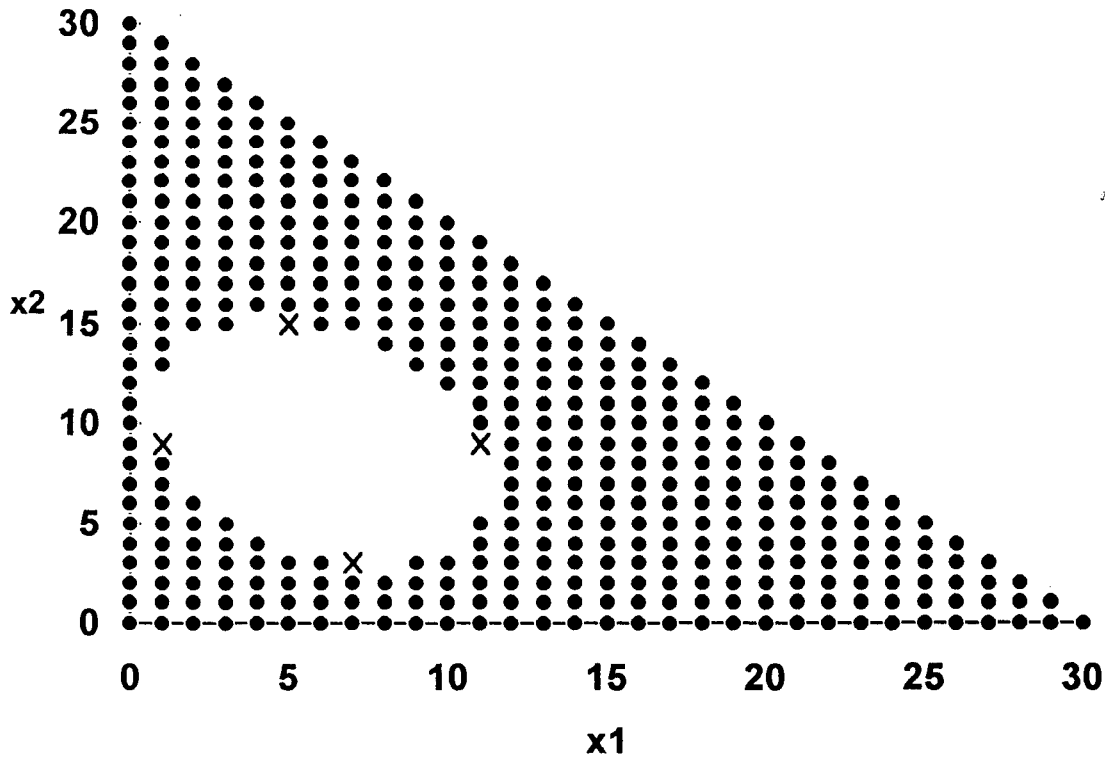
7.2.3. $\lambda = 0$, ponto aleatorizado (2 , 7 , 21), com probabilidade aproximada 0.659839



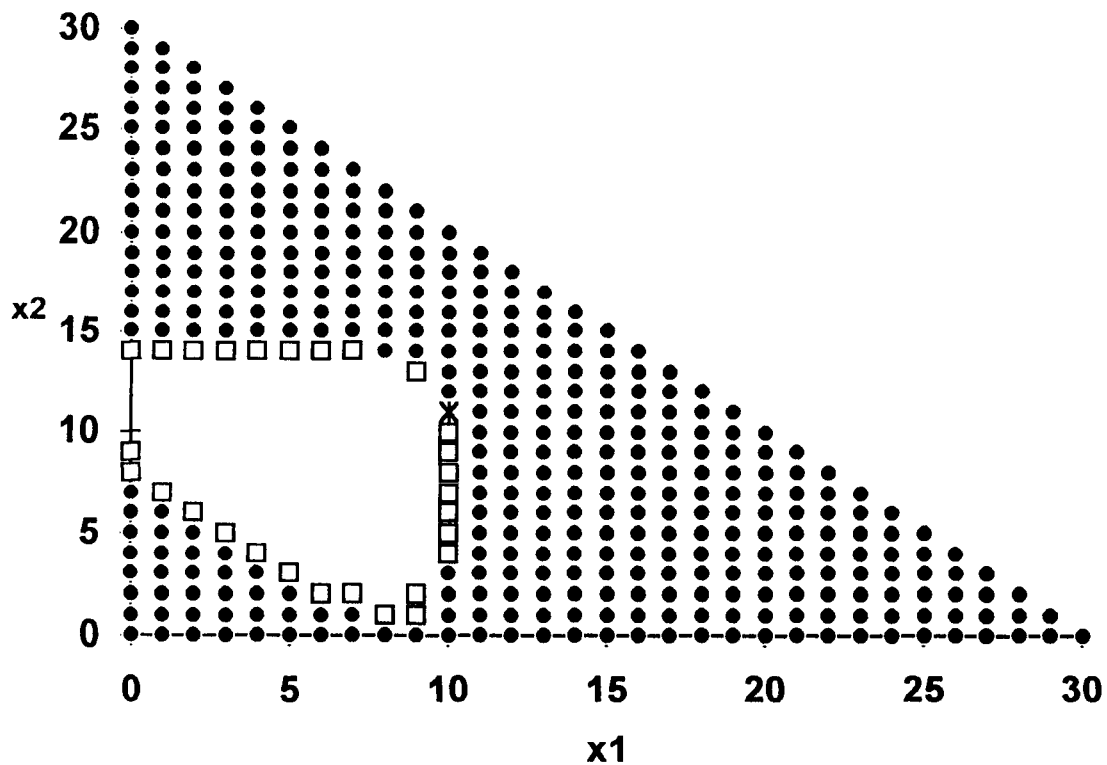
7.2.4. $\lambda = 2/3$, ponto aleatorizado (1 , 10 , 19), com probabilidade aproximada 0.593633



7.2.5. $\lambda = 1$, pontos aleatorizados $\{ (1, 9, 20), (7, 3, 20), (11, 9, 10), (5, 15, 10) \}$, com probabilidade aproximada 0.569679



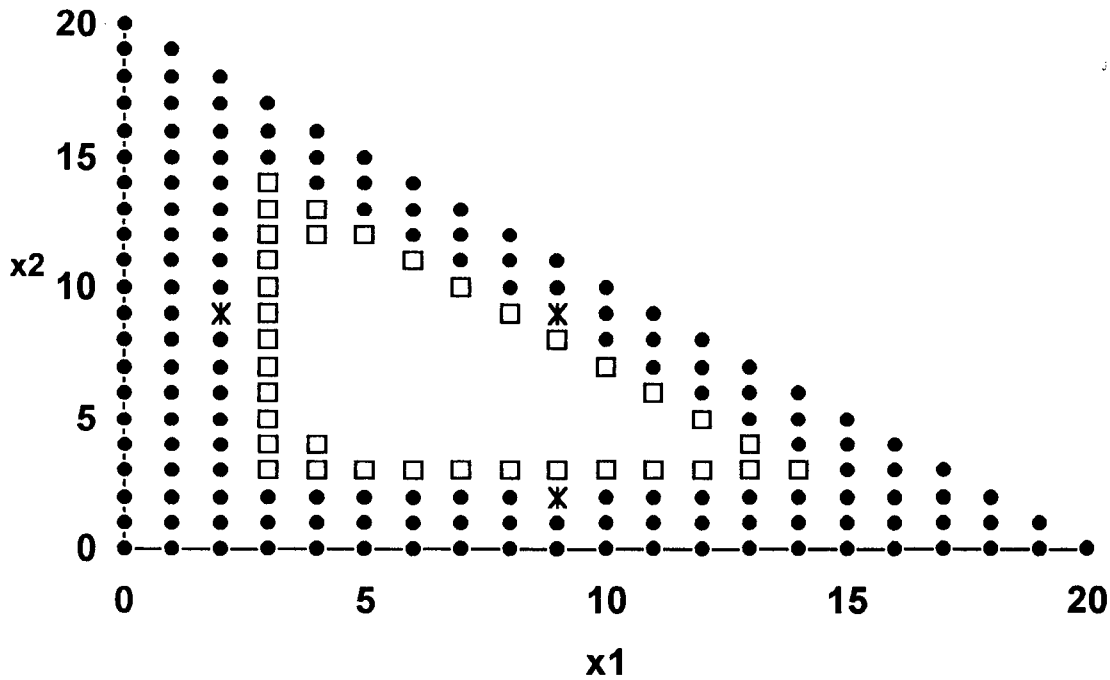
7.2.6. $\lambda = 5$, ponto aleatorizado $(10, 11, 9)$, com probabilidade aproximada 0.796866



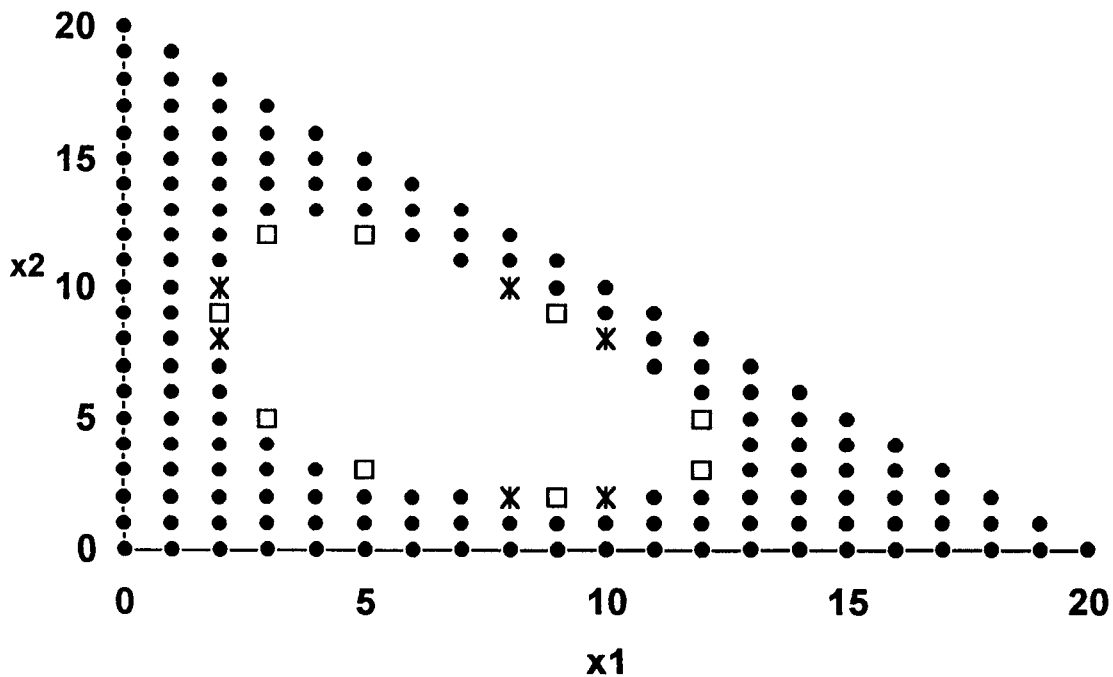
8. Hipótese nula dada por $(1/3, 1/3, 1/3)$.

8.1. Tamanho da amostra 20.

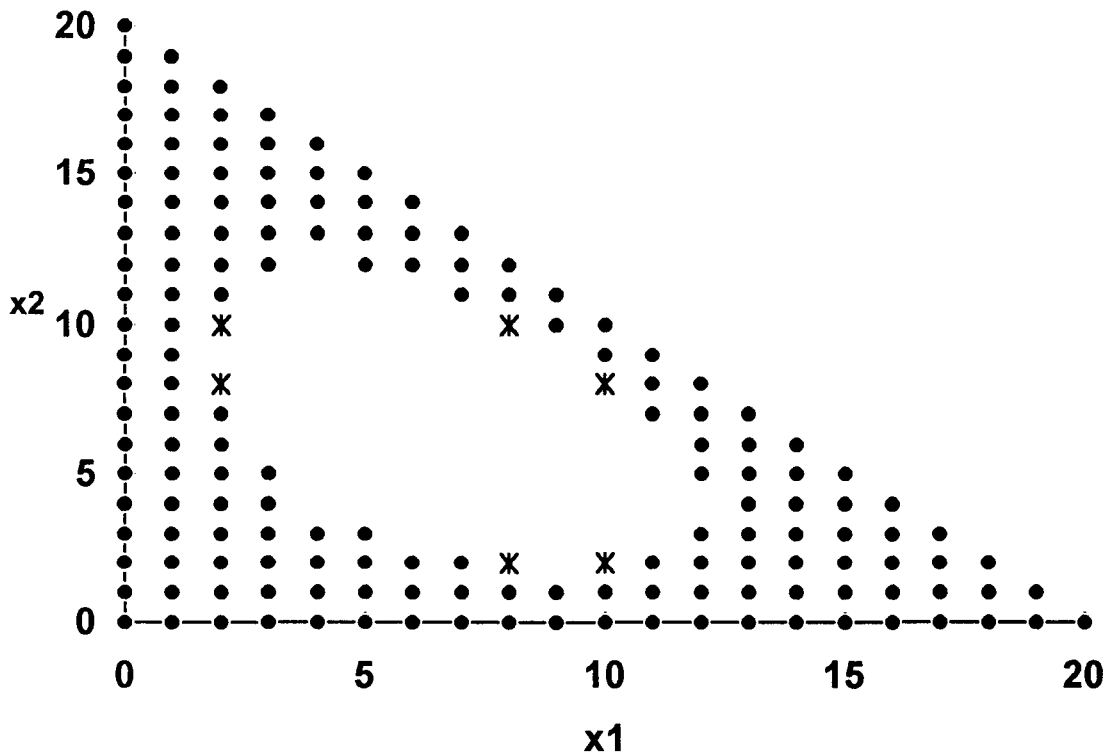
8.1.1. $\lambda = -5$, pontos aleatorizados $\{(2, 9, 9), (9, 2, 9), (9, 9, 2)\}$, com probabilidade aproximada 0.654473



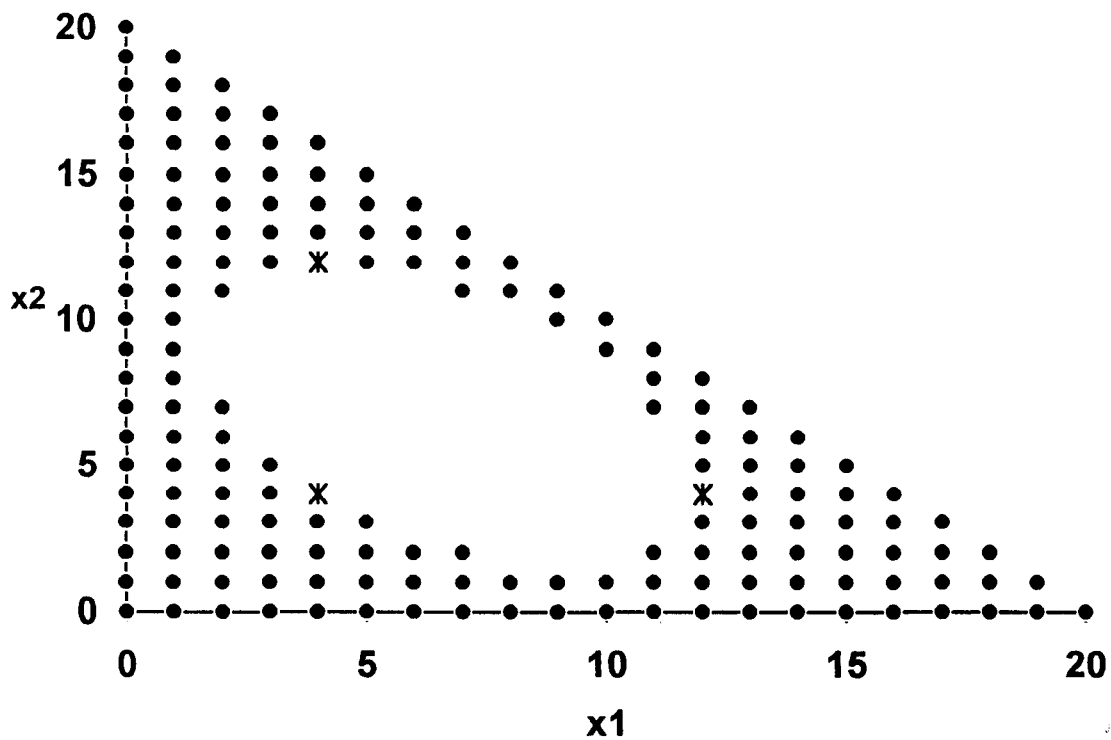
8.1.2. $\lambda = -1$, pontos aleatorizados $\{(2, 8, 10), (2, 10, 8), (8, 2, 10), (8, 10, 2), (10, 2, 8), (10, 8, 2)\}$, com probabilidade aproximada 0.990636



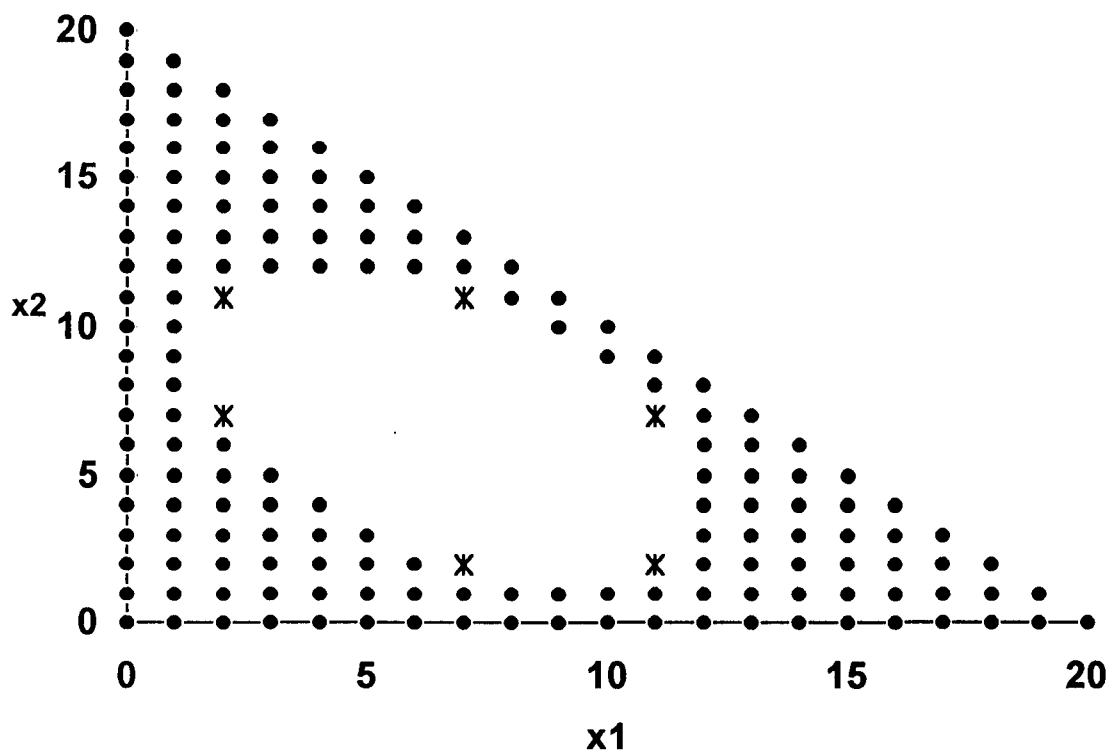
8.1.3. $\lambda = 0$, pontos aleatorizados $\{ (2, 8, 10), (2, 10, 8), (8, 2, 10), (8, 10, 2), (10, 2, 8), (10, 8, 2) \}$, com probabilidade aproximada 0.142151



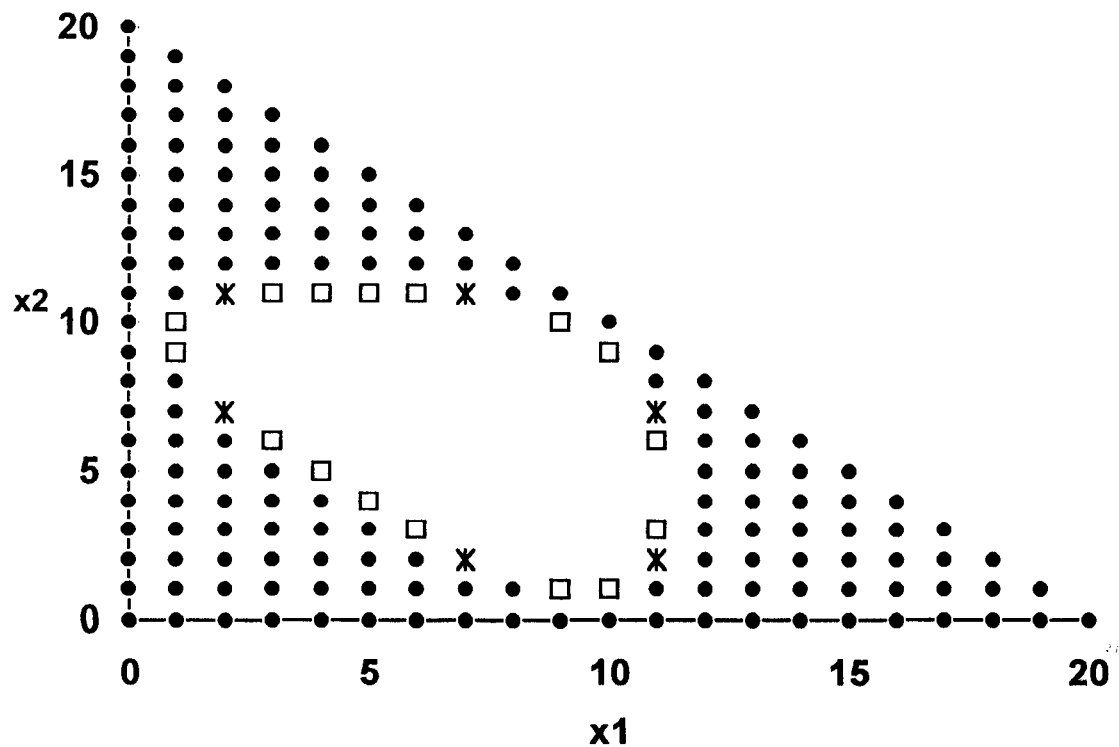
8.1.4. $\lambda = 2/3$, pontos aleatorizados $\{ (4, 4, 12), (4, 12, 4), (12, 4, 4) \}$, com probabilidade aproximada 0.268056



8.1.5. $\lambda = 1$, pontos aleatorizados $\{ (2, 7, 11), (2, 11, 7), (7, 2, 11), (7, 11, 2), (11, 2, 7), (11, 7, 2) \}$, com probabilidade aproximada 0.466291

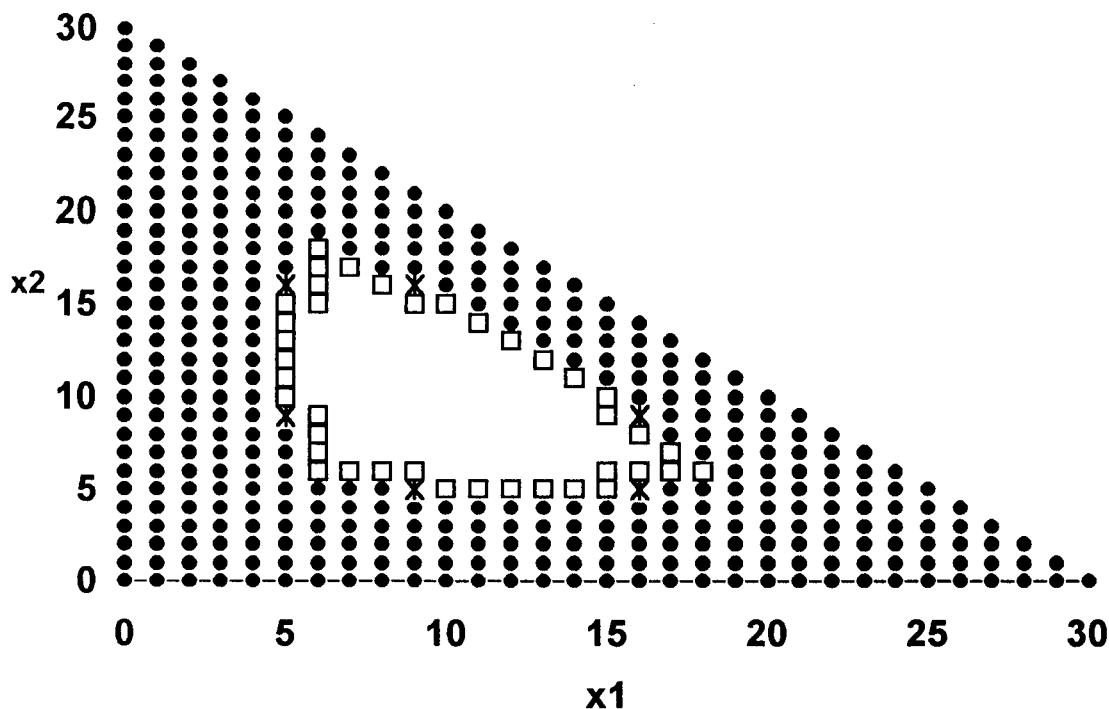


8.1.6. $\lambda = 5$, pontos aleatorizados $\{ (2, 7, 11), (2, 11, 7), (7, 2, 11), (7, 11, 2), (11, 2, 7), (11, 7, 2) \}$, com probabilidade aproximada 0.771846

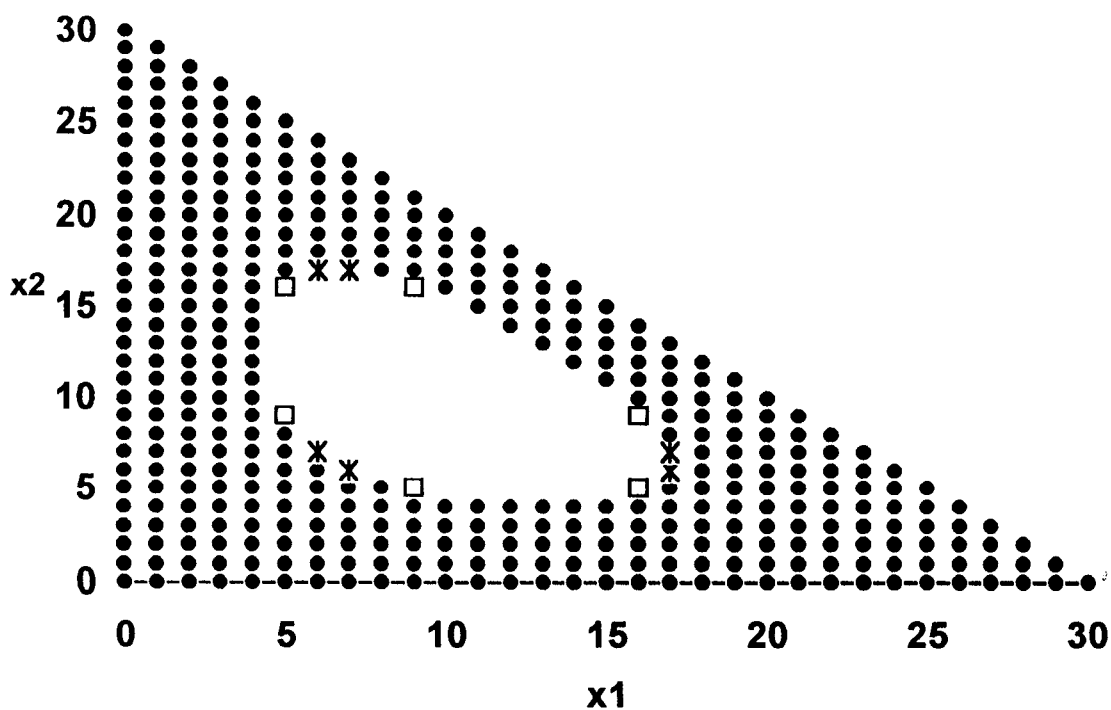


8.2. Tamanho da amostra 30.

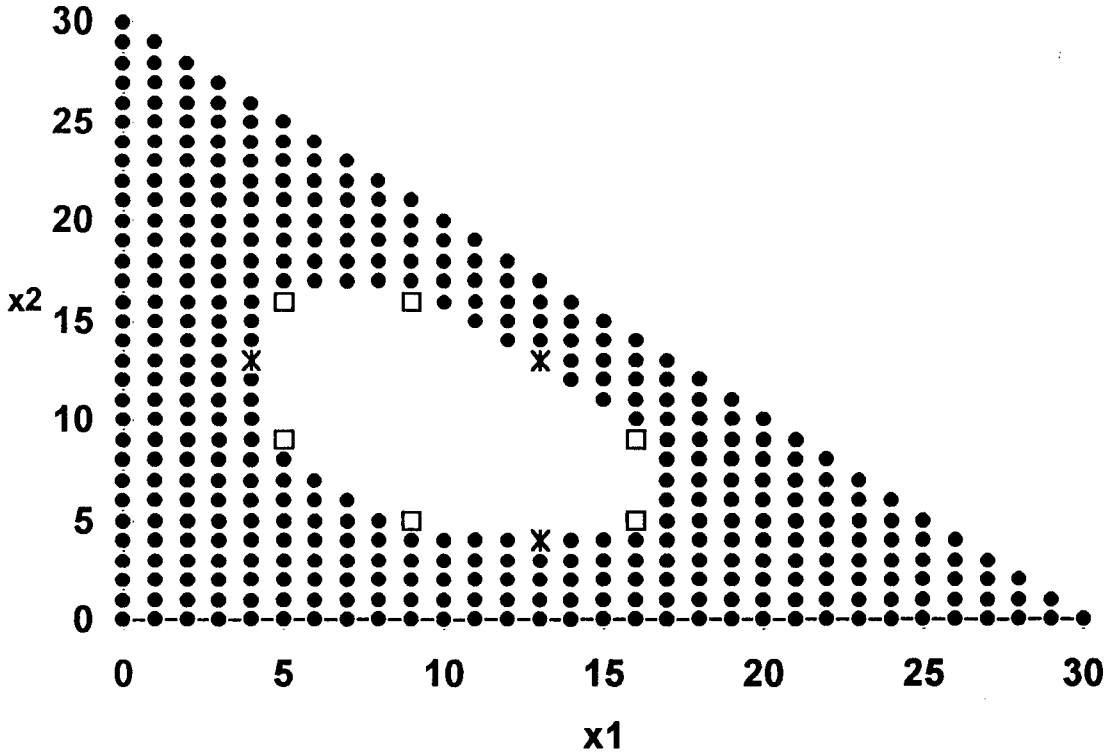
8.2.1. $\lambda = -5$, pontos aleatorizados $\{ (5, 9, 16), (5, 16, 9), (9, 5, 16), (9, 16, 5), (16, 5, 9), (16, 9, 5) \}$, com probabilidade aproximada 0.705511



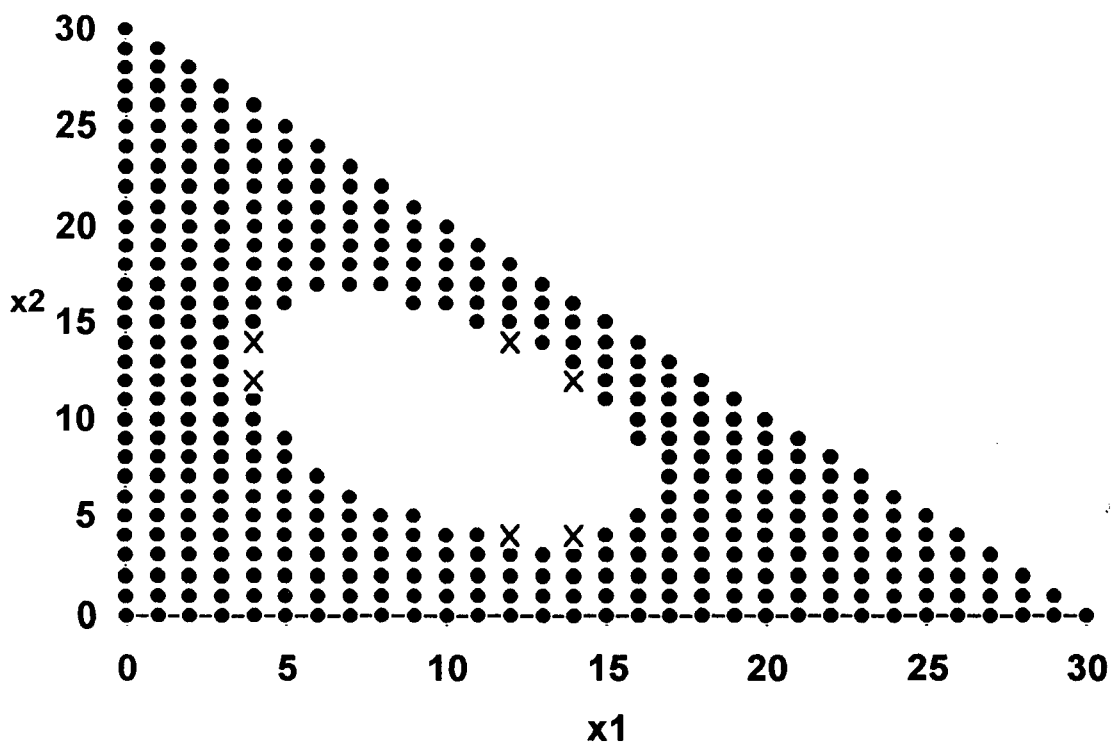
8.2.2. $\lambda = -1$, pontos aleatorizados $\{ (6, 7, 17), (6, 17, 7), (7, 6, 17), (7, 17, 6), (17, 6, 7), (17, 7, 6) \}$, com probabilidade aproximada 0.80503



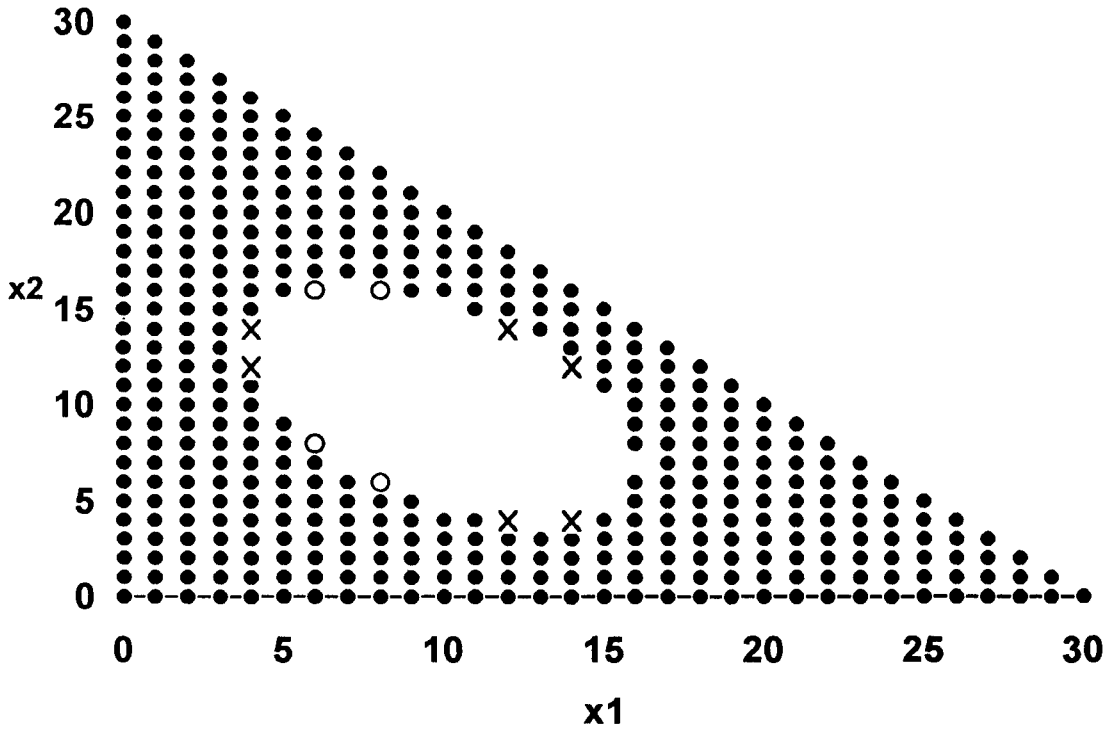
8.2.3. $\lambda = 0$, pontos aleatorizados $\{ (4, 13, 13), (13, 4, 13), (13, 13, 4) \}$, com probabilidade aproximada 0.71885



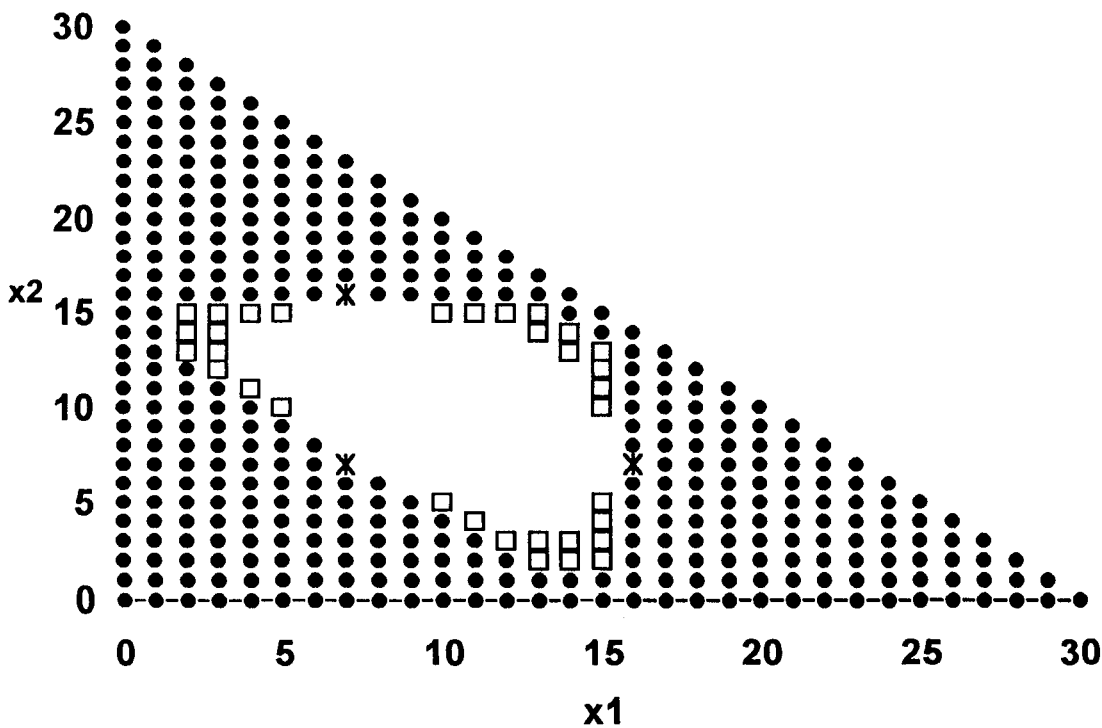
8.2.4. $\lambda = 2/3$, pontos aleatorizados $\{ (4, 12, 14), (4, 14, 12), (12, 4, 14), (12, 14, 4), (14, 4, 12), (14, 12, 4) \}$, com probabilidade aproximada 0.287073



8.2.5. $\lambda = 1$, pontos aleatorizados $\{ (4, 12, 14), (4, 14, 12), (12, 4, 14), (12, 14, 4), (14, 4, 12), (14, 12, 4), (6, 8, 16), (6, 16, 8), (8, 6, 16), (8, 16, 6), (16, 6, 8), (16, 8, 6) \}$, com probabilidade aproximada 0.108329



8.2.6. $\lambda = 5$, pontos aleatorizados $\{ (7, 7, 16), (7, 16, 7), (16, 7, 7) \}$, com probabilidade aproximada 0.112358



9. Análise dos casos particulares

Em uma análise preliminar das regiões de rejeição definidas pelas estatísticas pertencentes a família de estatísticas definidas por Cressie e Read faremos uma comparação entre os testes definidos em 4.1. pág. 39 e 4.2. pág. 40, observando-se a variação dos valores de λ nessa família, bem como a variação do tamanho da amostra. Estamos tratando com essa análise inicial a velocidade de convergência para a distribuição Qui-quadrado das estatísticas pertencentes a família de estatísticas definidas por Cressie e Read. Fazemos também uma comparação entre as regiões de rejeição para diferentes valores de λ , comparando-se assim algumas estatísticas da família de estatísticas definidas por Cressie e Read.

9.1. Hipótese nula dada por (0.1 , 0.2 , 0.7)

Sendo a hipótese nula dada por (0.1 , 0.2 , 0.7), para um tamanho de amostra igual a 30, podemos notar uma grande diferença entre as regiões de rejeição do teste de tamanho exato, aplicado em 4.2. pág. 40 e o teste Qui-quadrado, como aplicado em 4.1. pág. 39, para o valor de $\lambda = -5$, sendo essa diferença amenizada no caso do tamanho da amostra igual a 35, o que sugere que em um tamanho de amostra ainda maior possamos ter essa diferença ainda mais abrandada.

No caso de $\lambda = 5$, a diferença entre as regiões de rejeição dos testes não é tão grande como no caso $\lambda = -5$, porém apresenta ainda uma certa diferença. Dos 496 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 30, 422 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado, para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 16 pontos, o que representa um aumento indevido na região de rejeição em torno de 3.79%, obtendo-se um tamanho de teste de aproximadamente 0.1381 . Tratando-se do caso da amostra de tamanho 35, dos 666 pontos pertencentes ao espaço amostral, 574 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto

aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 22 pontos, o que representa um aumento indevido na região de rejeição em torno de 3.83%, obtendo-se um tamanho de teste de aproximadamente 0.1098, apresentando uma certa diferença do tamanho exato (0.05), assim como no caso da amostra de tamanho 30.

No caso de $\lambda = -1$, a diferença entre as regiões de rejeição dos testes fica menor do que no caso $\lambda = 5$, porém apresenta ainda uma leve diferença. Dos 496 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 30, 379 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 52 pontos, o que representa um aumento indevido na região de rejeição em torno de 13.72%, obtendo-se um tamanho de teste de aproximadamente 0.0831 . Tratando-se do caso da amostra de tamanho 35, dos 666 pontos pertencentes ao espaço amostral, 571 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 20 pontos, o que representa um aumento indevido na região de rejeição em torno de 3.5%, obtendo-se um tamanho de teste de aproximadamente 0.0726, um pouco mais próximo do tamanho exato (0.05) do que no caso da amostra de tamanho 30, que também é razoavelmente próximo.

Para os casos onde $\lambda \in \{ 0 , 2/3 , 1 \}$, temos uma diferença, das regiões de rejeição, muito pequena entre o teste de tamanho exato, como aplicado em 4.2. pág. 40, e o teste Qui-quadrado, como aplicado em 4.1. pág. 39, tanto para amostras de tamanho 30 como para amostras de tamanho 35. Além disso, podemos observar que as regiões, quando comparadas com diferentes valores de λ , apresentam uma grande semelhança, ou seja, os pontos de interseção para diferentes valores de λ , formam, com exceção de alguns pontos, as próprias regiões, tanto no caso de tamanho 30, e mais acentuado ainda no caso de tamanho 35.

9.2. Hipótese nula dada por (0.2 , 0.3 , 0.5)

Sendo a hipótese nula dada por (0.2 , 0.3 , 0.5), para um tamanho de amostra igual a 30, podemos notar uma certa diferença entre as regiões de rejeição do teste de tamanho exato, aplicado em 4.2. pág. 40 e o teste Qui-quadrado, como aplicado em 4.1. pág. 39, para o valor de $\lambda = -5$, sendo essa diferença muito pouco amenizada no caso do tamanho da amostra igual a 35. Dos 231 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 20, 138 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 61 pontos, o que representa um aumento indevido na região de rejeição em torno de 44.2%, obtendo-se um tamanho de teste de aproximadamente 0.2793 . Tratando-se do caso da amostra de tamanho 30, dos 496 pontos pertencentes ao espaço amostral, 344 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 91 pontos, o que representa um aumento indevido na região de rejeição em torno de 26.45%, obtendo-se um tamanho de teste de aproximadamente 0.2203, ainda bastante distante do tamanho exato (0.05) análogo ao caso da amostra de tamanho 20.

No caso de $\lambda = 5$, a diferença entre as regiões de rejeição dos testes não é tão grande como no caso $\lambda = -5$, porém apresenta ainda uma certa diferença. Dos 231 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 20, 154 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 30 pontos, o que representa um aumento indevido na região de rejeição em torno de 19.48%, obtendo-se um tamanho de teste de aproximadamente 0.1583 . Tratando-se do caso da amostra de tamanho 30, dos 496 pontos pertencentes ao espaço amostral, 380 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 28 pontos, o que representa um aumento indevido na região de rejeição em torno de 7.37%, obtendo-se um tamanho de teste de aproximadamente 0.1108, um pouco mais próximo do

tamanho exato (0.05) do que no caso da amostra de tamanho 20, porém apresentando ainda uma certa diferença.

Para os casos onde $\lambda \in \{ -1 , 0 , 2/3 , 1 \}$, temos uma pequena diferença das regiões de rejeição entre o teste de tamanho exato, como aplicado em 4.2. pág. 40, e o teste Qui-quadrado, como aplicado em 4.1. pág. 39, tanto para amostra de tamanho 20 como para amostra de tamanho 30. Além disso, podemos observar que as regiões, quando comparadas com diferentes valores de λ , apresentam uma grande semelhança, ou seja, os pontos de interseção para diferentes valores de λ , formam, com exceção de alguns pontos, as próprias regiões, tanto no caso de tamanho 20, e mais acentuado ainda no caso de tamanho 30.

9.3. Hipótese nula dada por (1/3 , 1/3 , 1/3)

Sendo a hipótese nula dada por (1/3 , 1/3 , 1/3), para um tamanho de amostra igual a 20, podemos notar uma certa diferença das regiões de rejeição entre o teste de tamanho exato, aplicado em 4.2. pág. 40 e o teste Qui-quadrado, como aplicado em 4.1. pág. 39, para o valor de $\lambda = -5$, sendo essa diferença um pouco amenizada no caso do tamanho da amostra igual a 30. Dos 231 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 20, 153 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 35 pontos, o que representa um aumento indevido na região de rejeição em torno de 22.88%, obtendo-se um tamanho de teste de aproximadamente 0.1879. Tratando-se do caso da amostra de tamanho 30, dos 496 pontos pertencentes ao espaço amostral, 387 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 39 pontos, o que representa um aumento indevido na região de rejeição em torno de 10.08%, obtendo-se um tamanho de teste de aproximadamente 0.1487, ainda distante do tamanho exato (0.05), porém um pouco menos distante do que no caso da amostra de tamanho 20.

No caso de $\lambda = 5$, a diferença entre as regiões de rejeição dos testes não é tão grande como no caso $\lambda = -5$, porém apresenta ainda uma certa diferença. Dos 231 pontos pertencentes ao espaço amostral, tratando-se do caso da amostra de tamanho 20, 156 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 16 pontos, o que representa um aumento indevido na região de rejeição em torno de 10.26%, obtendo-se um tamanho de teste de aproximadamente 0.1162. Tratando-se do caso da amostra de tamanho 30, dos 496 pontos pertencentes ao espaço amostral, 369 pertencem a região de rejeição para o teste de tamanho exato, como aplicado em 4.2. pág. 40, incluindo-se a estes o ponto aleatorizado; para o teste Qui-quadrado, como aplicado em 4.1. pág. 39, incluímos além dos citados anteriormente mais 33 pontos, o que representa um aumento indevido na região de rejeição em torno de 8.94%, obtendo-se um tamanho de teste de aproximadamente 0.0814, um bem mais próximo do tamanho exato (0.05) do que no caso da amostra de tamanho 20.

Para os casos onde $\lambda \in \{ -1 , 0 , 2/3 , 1 \}$, temos uma diferença muito pequena das regiões de rejeição entre o teste de tamanho exato, como aplicado em 4.2. pág. 40, e o teste Qui-quadrado, como aplicado em 4.1., tanto para amostra de tamanho 20 como para amostra de tamanho 30, onde em alguns casos este é menor do que aquele e em outros casos o inverso. Além disso, podemos observar que as regiões, quando comparadas com diferentes valores de λ , apresentam uma grande semelhança, ou seja, os pontos de interseção para diferentes valores de λ , formam, com exceção de alguns pontos, as próprias regiões, tanto no caso de tamanho 20, e mais acentuado ainda no caso de tamanho 30.

CAPÍTULO IV

CONCLUSÕES

1. Aplicação da Teoria da Informação na Estatística

O estudo da Teoria da Informação é fundamentado no conceito de mensagem, associados aos elementos de sua geração, transmissão e captação, onde existem falhas (diferenças) entre a mensagem inicial e a mensagem final recebida. Analisando essas falhas como fenômenos casuísticos, podemos abordar a mensagem recebida como aleatória, sendo governada sob alguma lei probabilística. Podemos associar então uma medida de flutuação na transmissão da mensagem, essa medida de flutuação, chamada de Entropia, foi tratada em Shannon, C. E. (1948) [16] e em seguida por muitos outros autores.

Fazendo-se um paralelo com a Estatística, no caso de variáveis aleatórias com distribuição multinomial, temos diferentes flutuações dos valores amostrados em repetições de realizações de variáveis aleatórias com diferentes distribuições multinomiais, ou seja, sendo (X_1, X_2, \dots, X_k) uma variável aleatória (vetor), com distribuição multinomial de parâmetros n (n inteiro positivo, fixo) e $(\theta_1, \theta_2, \dots, \theta_k)$, $((\theta_1, \theta_2, \dots, \theta_k) \in \Delta)$, tomando-se repetições de realizações de (X_1, X_2, \dots, X_k) , para cada valor possível de $(\theta_1, \theta_2, \dots, \theta_k)$, notamos diferentes flutuações no que é gerado. Para alguns valores de $(\theta_1, \theta_2, \dots, \theta_k)$, caso sejam observados algumas gerações de (X_1, X_2, \dots, X_k) , notamos grandes frequências em poucos pontos do espaço amostral (pouca variação de observações) e em outros valores de $(\theta_1, \theta_2, \dots, \theta_k)$ notamos muitos pontos com pouca frequência (muita variação de observações), o que nos leva a uma maior flutuação dos valores amostrados em muitos pontos com pouca frequência.

Utilizando o conceito de Entropia, como definido em Shannon, C. E. (1948) [16], podemos medir o grau de flutuação apresentado na realização de variáveis aleatórias com distribuição multinomial, sendo esse grau de flutuação maior, quanto mais próximo estiver de $(1/k, 1/k, \dots, 1/k)$ o valor do parâmetro $(\theta_1, \theta_2, \dots, \theta_k)$ e menor, quanto mais próximo estiver $(\theta_1, \theta_2, \dots, \theta_k)$ de um vértice do espaço paramétrico, por exemplo $(1, 0, \dots, 0)$.

2. Divergência Direcionada

Partindo-se do conceito de Entropia, definido em Shannon, C. E. (1948) [16] e sua aplicabilidade na Estatística, podemos definir uma medida de afinidade entre funções de probabilidade de variáveis aleatórias discretas. Uma dessas medidas, a divergência direcionada, pôde ser vista em Kullback, S. (1959) [7], sendo esta generalizadas por outros autores.

A Divergência Direcionada, aplicada em distribuições multinomiais, pode ser vista como uma medida de afinidade entre um ponto do espaço paramétrico Δ e os outros pontos desse espaço. Temos nessa medida um certo caráter direcional, pois a divergência direcionada de um certo ponto $\theta(1)$ em relação a outro ponto $\theta(2)$, pertencentes ao mesmo espaço paramétrico Δ é normalmente diferente da divergência direcionada de $\theta(2)$ em relação a outro ponto $\theta(1)$, podemos denotar esse caráter diferencial defido ao fato de que quando tomamos uma certa variável, desse modelo, como sendo uma outra, nem sempre é tão drástico como o contrário.

Tomando-se como base a bibliografia citada, bem como a semelhança existente entre essas duas divergências direcionadas, podemos dizer que possivelmente, com base na Divergência Direcionada Generalizada de ordem α proposta por Rathie e Kannappan, Cressie, N. A. C. e Read, T. R. C. (1984) [5] construíram uma Divergência Direcionada Generalizada.

2.1. Divergência Direcionada em Testes de Hipóteses

Utilizando o conceito de Divergência Direcionada Generalizada, pode-se encarar um teste de hipóteses multinomial como uma aplicação dessa divergência, onde ao invés de medir o grau de afinidade entre uma certa distribuição multinomial e outras possíveis, dentro do mesmo espaço paramétrico, é medido a plausibilidade de um certo valor observado, na geração de uma variável aleatória com distribuição multinomial, e possíveis valores dos parâmetros dessa distribuição, geradora de tais dados.

A plausibilidade de um certo valor observado, pertencente ao espaço amostral, com possíveis valores dos parâmetros de uma distribuição multinomial, geradora de tais dados, pode ser denotada pela divergência existente entre esse valor observado e o valor esperado na geração dessa variável aleatória com distribuição multinomial.

Com base nessas idéias estamos ligando a família de estatísticas definida em Cressie, N. A. C. e Read, T. R. C. (1984) [5] à teoria da informação.

Em Cressie, N. A. C. e Read, T. R. C. (1984) [5] temos a generalização de uma certa quantidade de testes de adequabilidade de ajuste em distribuições multinomiais, observados como casos particulares de uma função de divergência, indexada por um parâmetro λ , onde em certos casos particulares desta, ou seja, para alguns específicos λ 's podemos encontrar testes como o definido por K. Pearson e o Logaritmo da Razão de Verossimilhança.

3. Comparação de Testes

Em Cochran, W. G. (1952) [3], após um minucioso estudo da estatística χ^2 apresentada por K. Pearson, temos a sugestão de uma outra estatística, também com convergência para a distribuição Qui-quadrado, essa estatística é dada pelo logaritmo da razão de verossimilhança, iniciando-se assim uma extensa discussão entre qual dessas duas estatísticas, a χ^2 de Pearson e a

estatística do logaritmo da razão de verossimilhança, é a mais bem empregada em testes de adequabilidade de ajuste em distribuições multinomiais.

Com o trabalho de Cressie, N. A. C. e Read, T. R. C. (1984) [5], temos uma grande ferramenta para comparação entre, não somente essas duas estatísticas, a χ^2 de Pearson e a estatística do logaritmo da razão de verossimilhança, porém algumas outras que surgiram nessa longa discussão, tais como: estatística X^2 de Neymann, estatística do logaritmo da razão de verossimilhança modificado, estatística de Freeman - Tukey, pois estas ficam enquadradas como casos particulares da função de divergência de Cressie e Read.

3.1. Comparação Assintótica

Podemos ver em Cressie, N. A. C. e Read, T. R. C. (1984) [5] que em relação à eficiência assintótica, os testes utilizando as estatísticas da família $2nI^\lambda$, para o caso do número de células k fixo, mostram-se sem grandes diferenças, porém em relação à velocidade de convergência para a distribuição Qui-quadrado existem valores de λ que apresentam uma convergência mais rápida, no caso, para $\lambda \in [0.2 , 2.7]$, o que inclui as estatísticas de K. Pearson ($\lambda = 1$) e a estatística definida por Cressie e Read ($\lambda = 2/3$), porém exclui a estatística do logaritmo da razão de verossimilhança ($\lambda = 0$).

Temos então através de Cressie, N. A. C. e Read, T. R. C. (1984) [5] que as estatísticas da família família $2nI^\lambda$, para o caso do número de células k fixo, mostram-se semelhantes para os valores de λ pertencentes a $[0.2 , 2.7]$.

3.2. Comparação em Amostras Finitas

No presente trabalho foram feitas simulações de testes de adequabilidade de ajuste em distribuições multinomiais para os seguintes casos particulares:

Hipótese nula	valor de λ	tamanhos de amostra
(0.1, 0.2, 0.7)	-5	30 e 35
	-1	30 e 35
	0	30 e 35
	2/3	30 e 35
	1	30 e 35
	5	30 e 35
(0.2, 0.3, 0.5)	-5	20 e 30
	-1	20 e 30
	0	20 e 30
	2/3	20 e 30
	1	20 e 30
	5	20 e 30
(1/3, 1/3, 1/3)	-5	20 e 30
	-1	20 e 30
	0	20 e 30
	2/3	20 e 30
	1	20 e 30
	5	20 e 30

Podemos ver através dos gráficos mostrados no capítulo III que as áreas de rejeição no teste de tamanho exato definido em 4.2. pág. 40 e no teste Qui-quadrado definido em 4.1. pág. 39, para

diferentes valores de λ , nos casos $\lambda \in \{ 0, 2/3, 1 \}$, para o caso da hipótese nula simples dada por $(0.1, 0.2, 0.7)$, possuem uma diferença muito pequena, para os tamanhos de amostra 30 e 35. No caso da hipótese nula simples dada por $(0.2, 0.3, 0.5)$ ou $(1/3, 1/3, 1/3)$ temos as áreas de rejeição no teste de tamanho exato definido em 4.2. pág. 40 e no teste Qui-quadrado definido em 4.1. pág. 39, além dos citados valores de λ ($\lambda \in \{ 0, 2/3, 1 \}$), $\lambda = -1$, apresentando pequenas diferenças, para os tamanhos de amostra 20 e 30; o que pode nos levar a um questionamento sobre a velocidade de convergência comentada em Cressie, N. A. C. e Read, T. R. C. (1984) [5]

Observando-se os pontos de interseção nas regiões de rejeição para os valores de $\lambda \in \{ 0, 2/3, 1 \}$ nos respectivos testes exatos definido em 4.2. pág. 40, levando-se em consideração a aleatorização definida em 4.3. pág. 40, temos as seguintes probabilidades de uma realização da respectiva variável aleatória, sob a hipótese nula considerada como verdadeira, pertencer a essas regiões de interseção.

i1) Hipótese nula $(0.1, 0.2, 0.7)$, $n = 30$. Probabilidade $\cong 0.0267149$

i2) Hipótese nula $(0.1, 0.2, 0.7)$, $n = 35$. Probabilidade $\cong 0.0267223$

ii1) Hipótese nula $(0.2, 0.3, 0.5)$, $n = 20$. Probabilidade $\cong 0.0472493$

ii2) Hipótese nula $(0.2, 0.3, 0.5)$, $n = 30$. Probabilidade $\cong 0.04144$

iii1) Hipótese nula $(1/3, 1/3, 1/3)$, $n = 20$. Probabilidade $\cong 0.0424132$

iii2) Hipótese nula $(1/3, 1/3, 1/3)$, $n = 30$. Probabilidade $\cong 0.0437907$

Temos então que as estatísticas: χ^2 de Pearson, logaritmo da razão de verossimilhança e estatística da família Cressie e Read para $\lambda = 2/3$, para os casos das hipóteses nulas pertencentes a $\{ (0.2, 0.3, 0.5), (1/3, 1/3, 1/3) \}$, nos casos de amostras de tamanho 20 e 30, e para a hipótese nula $(0.1, 0.2, 0.7)$, nos casos de amostras de tamanho 30 e 35, apresentam regiões de rejeição muito semelhantes.

Baseado no que foi exposto até então, podemos dizer que a família de estatísticas generalizadas por intermédio da função de divergência apresentada em Cressie, N. A. C. e Read, T. R. C. (1984) [5], produzem resultados bastante semelhantes no tocante à aplicação destas estatísticas em testes de adequabilidade de ajuste em distribuições multinomiais, para um certo conjunto de estatísticas pertencentes a esta família de estatísticas, onde em Cressie, N. A. C. e Read, T. R. C. (1984) [5] é comentado que seja para valores de λ pertencentes a $[0.2, 2.7]$ e neste trabalho ampliamos para pelo menos a inclusão neste intervalo de valores até o valor zero, ou seja, teríamos valores de λ pertencentes a $[0, 2.7]$.

4. Conclusões Finais

Um dos objetivos desse trabalho era o de ligar a família de estatísticas apresentadas em Cressie, N. A. C. e Read, T. R. C. (1984) [5] a alguma teoria já existente, para não ficarmos com a impressão de que essa família de estatísticas tinham surgido do além. Essa ligação foi realizada com a Teoria da Informação, devido principalmente a bibliografia citada no próprio artigo de Cressie e Read. Expomos alguns tópicos da Teoria da Informação e construímos passo a passo um possível trajeto para a obtenção dessa família de estatísticas, onde esse trajeto em momento algum é exposto explicitamente no artigo acima citado nem tão pouco em Read, T. R. C. e Cressie, N. A. C. (1988) [14].

O outro objetivo que provocou o surgimento desse trabalho foi o questionamento da semelhança existente entre as estatísticas: χ^2 de Pearson e logaritmo da razão de verossimilhança, para aplicações de testes de adequabilidade de ajuste em distribuições multinomiais.

Para analisar a semelhança existente entre as estatísticas utilizamos a família de estatísticas definidas em Cressie, N. A. C. e Read, T. R. C. (1984) [5], através de uma simulação de alguns casos particulares de estatísticas pertencentes a essa família, em aplicações para testes de adequabilidade de ajuste em distribuições multinomiais e apresentamos por intermédio de gráficos que as regiões de rejeição para esses casos particulares. Para algumas das estatísticas utilizadas essas regiões de rejeição apresentavam uma razoável semelhança. Entre as estatísticas, cujas regiões de rejeição apresentavam uma razoável semelhança, estavam as estatísticas: χ^2 de Pearson, logaritmo da razão de verossimilhança e a estatística sugerida por Cressie e Read ($\lambda = 2/3$).

Temos então, baseado no que foi exposto até agora, que as estatísticas : χ^2 de Pearson e logaritmo da razão de verossimilhança, quando na aplicação em testes de adequabilidade de ajuste, apresentam uma considerável semelhança, bem como um certo subconjunto de estatísticas pertencentes a família de estatísticas de Cressie e Read.

APÊNDICE 1

Programa para formar as regiões de rejeição aleatorizada para a família de estatísticas de Cressie e Read, com tamanho real α (Pacote Mathematica)

(* Teste alternativo de ajustamento de distribuição, utilizando o conceito de Cressie & Read *)

(* (v1 , v2 , v3) : Hipótese nula; q : Valor de λ na família de estatísticas de Cressie e Read;

n : tamanho da amostra; alfa : tamanho do teste *)

Clear [hip , v1 , v2 , v3 , q , n , alfa];

hip [v1_ , v2_ , v3_ , q_ , n_ , alfa_] := (

Clear [ln , ff , f , l , lp , i , j , lcr , val , soma , k , t , aux , beta , poder];

ln [x_] := If [x == 0 , 0 , Log [x]];

ff [x_ , y_] := Multinomial [x , y , n - x - y] v1 ^ x v2 ^ y v3 ^ (n - x - y);

f [x_ , y_] := If [q == 0 ,

2 (x ln [x / (n v1)] + y ln [y / (n v2)] + (n - x - y) ln [(n - x - y) / (n (1 - v1 - v2))]),

If [q == -1 , If [x y == 0 || x + y == n , Infinity ,

2 (n v1 Log [n v1 / x] + n v2 Log [n v2 / y] + n (1 - v1 - v2) Log [n (1 - v1 - v2) / (n - x - y)

)]],

If [q == 99 , 1 / ff [x , y] , If [q > 0 ,

2 / (q (q + 1)) (x (x / (n v1)) ^ q + y (y / (n v2)) ^ q + (n - x - y) ((n - x - y) / (n (1 - v1 - v2))) ^ q - n) ,

If [x y == 0 || x + y == n , Infinity ,

2 / (q (q + 1)) (x (x / (n v1)) ^ q + y (y / (n v2)) ^ q + (n - x - y) ((n - x - y) / (n (1 - v1 - v2))) ^ q - n)]]]]];

l = Table [{ x , y , f [x , y] } , { x , 0 , n } , { y , 0 , n - x }];

```

lp = { };
Do [ lp = Join [ lp , l [ [ i ] ] ] , { i , n + 1 } ];
Print [ " Ordenando os dados " ];
Do [ Do [ If [ lp [ [ i , 3 ] ] < lp [ [ j , 3 ] ] , x = lp [ [ i ] ] ; lp [ [ i ] ] = lp [ [ j ] ] ; lp [ [ j ] ] = x ]
, { j , i + 1 , ( n + 1 ) ( n + 2 ) / 2 } ] , { i , 1 , ( n + 1 ) ( n + 2 ) / 2 - 1 } ];
soma = 0;
k = 0;
Print [ " Calculando o valor critico do teste " ];
While [ soma < alfa , k = k + 1 ; soma = soma + ff [ lp [ [ k , 1 ] ] , lp [ [ k , 2 ] ] ] ];
j = k - 1;
i = k + 1;
While [ lp [ [ j , 3 ] ] == lp [ [ k , 3 ] ] , j = j - 1 ];
While [ lp [ [ i , 3 ] ] == lp [ [ k , 3 ] ] , i = i + 1 ];
val = { };
Do [ val = Join [ val , { lp [ [ t ] ] } ] , { t , j + 1 , i - 1 } ];
lcr = { };
Do [ lcr = Join [ lcr , { { lp [ [ t , 1 ] ] , lp [ [ t , 2 ] ] } } ] , { t , 1 , i - 1 } ];
ListPlot [ lcr ];
If [ i > k + 1 , Do [ soma = soma + ff [ lp [ [ t , 1 ] ] , lp [ [ t , 2 ] ] ] , { t , k + 1 , i - 1 } ] ];
aux = Sum [ ff [ lp [ [ t , 1 ] ] , lp [ [ t , 2 ] ] ] , { t , j + 1 , i - 1 } ];
beta = ( alfa - soma + aux ) / aux;

Print [ " Fim do programa; para saber o poder do teste digite : poder ( p1 , p2 , p3 ) , ( p1 , p2 , p3 )

```

```
no simplex " ];
```

```
Print [ " " ];
```

```
(* funcao para medir o poder do teste para a regioao da hipotese alternativa *)
```

```
poder [ b1_, b2_, b3_ ] := (
```

```
Clear [ g , pd ];
```

```
g [ x_ , y_ ] := Multinomial [ x , y , n - x - y ] b1 ^ x b2 ^ y b3 ^ ( n - x - y );
```

```
pd = 0;
```

```
Do [ Do [ If [ f [ x , y ] > lp [ [ k , 3 ] ] , pd = pd + g [ x , y ] , If [ f [ x , y ] == lp [ [ k , 3 ] ] , pd =
```

```
pd + beta g [ x , y ] ] ]
```

```
, { y , 0 , n - x } ] , { x , 0 , n } ];
```

```
Print [ " " ] ; Print [ " Poder [ " , b1 , " , " , b2 , " , " , b3 , " ] = " , pd ] ) )
```

APÊNDICE 2

Programa para formar as regiões de rejeição para a família de estatísticas de Cressie e Read, utilizando o conceito tradicional de testes Qui-quadrado. (Pacote Mathematica)

(* Teste alternativo de ajustamento de distribuicao, utilizando o conceito de Cressie & Read *)

(* (v1 , v2 , v3) : Hipótese nula; q : Valor de λ na família de estatísticas de Cressie e Read;

n : tamanho da amostra *)

Clear[hip , v1 , v2 , v3 , q , n];

hip[v1_ , v2_ , v3_ , q_ , n_] := (

Clear[ln , ff , f , l , lp , rr , soma];

ln[x_] := If[x == 0 , 0 , Log [x]];

ff[x_ , y_] := Multinomial [x , y , n-x-y] v1 ^ x v2 ^ y v3 ^ (n - x - y);

f[x_ , y_] := If[q == 0 ,

2 (x ln [x / (n v1)] + y ln [y / (n v2)] + (n - x - y) ln [(n - x - y) / (n (1 - v1 - v2))]) ,

If[q == -1 , If[x y == 0 || x + y == n , Infinity,

2 (n v1 Log [n v1 / x] + n v2 Log [n v2 / y] + n (1 - v1 - v2) Log [n (1 - v1 - v2) / (n - x - y)

)] ,

If[q == 99 , 1 / ff[x , y] , If[q > 0 ,

2 / (q (q + 1)) (x (x / (n v1)) ^ q + y (y / (n v2)) ^ q + (n - x - y) ((n - x - y) / (n (1 - v1 - v2))) ^ q - n) ,

If[x y == 0 || x + y == n , Infinity,

2 / (q (q + 1)) (x (x / (n v1)) ^ q + y (y / (n v2)) ^ q + (n - x - y) ((n - x - y) / (n (1 - v1 - v2))) ^ q - n)]]]]];

l = Table [{ x , y } , { x , 0 , n } , { y , 0 , n - x }];

```
lp = { };  
Do [ lp = Join [ lp , l [ [ i ] ] ], { i , n + 1 } ];  
rr = { };  
Do [ If [ f [ lp [ [ i , 1 ] ] , lp [ [ i , 2 ] ] ] > 5.99146454710798 , rr = Join [ rr , { lp [ [ i ] ] } ] ], { i ,  
Length [ lp ] } ];  
soma = 0;  
Do [ soma = soma + ff [ rr [ [ i , 1 ] ] , rr [ [ i , 2 ] ] ] , { i , Length [ rr ] } ];  
ListPlot [ rr ];  
{ " hip " , { v1 , v2 , v3 } , " labda " , q , " n " , n , " tamanho da amostra " , soma } )
```

BIBLIOGRAFIA

- [1] Ash, R. (1965)
Information Theory.
John Wiley & Sons.
- [2] Bickel, P. J. e Doksum, K. A. (1977)
Mathematical Statistics.
Prentice Hall.
- [3] Cochran, W. G. (1952)
The χ^2 Test of Goodness of Fit
The Annals of Mathematical Statistics. 23, 315 - 345.
- [4] Cohen, A. e Sackrowitz, H. B. (1975)
Unbiasedness of The Chi-square, Likelihood Ratio, and Other Goodness of Fit Tests For The Equal Cell Case.
The Annals of Statistics. 3, 959 - 964.
- [5] Cressie, N. A. C. e Read, T. R. C. (1984)
Multinomial Goodness - of - fit Tests.
Journal of The Royal Statistical Society B. 46, 440 - 464.
- [6] Kannappan, P. (1972)
On Shannon's Entropy, Directed Divergence and Inaccuracy.
Zeitschrift für Wahr. Und Verwandte Gebiete. 22, 95 - 100.
- [7] Kullback, S. (1959)
Information Theory and Statistics.
John Wiley.
- [8] Kullback, S. e Leibler, R. A. (1951)
On Information and Sufficiency.
The Annals of Mathematical Statistics. 22, 79 - 86.
- [9] Lancaster, H. O. (1969)
The Chi-Squared Distribution
John Wiley & Sons.
- [10] Mathai, A. M. e Rathie, P. N. (1975)
Basic Concepts in Information Theory and Statistics.
John Wiley.
- [11] McMillan, B. (1953)
The Basic Theorems of Information Theory.
The Annals of Mathematical Statistics. 24, 196 - 219.

- [12] Radlow, R. e Alf Jr., E. F. (1975)
An Alternate Multinomial Assessment of The Accuracy of the χ^2 Test of Goodness of Fit.
Journal of the American Statistical Association. 70, 811 - 813.

- [13] Rathie, P. N. e Kannappan, P. (1972)
A Directed - Divergence Function of Type β .
Information and Control. 20, 38 - 45.

- [14] Read, T. R. C. e Cressie, N. A. C. (1988)
Goodness - of - fit Statistics For Discrete Multivariate Data.
Springer - Verlag

- [15] Rényi, A. (1970)
Probability Theory.
North - Holland.

- [16] Shannon C. E. (1948)
A Mathematical Theory of Communication.
The Bell System Technical Journal. 27, 379 - 423.

- [17] Pearson, K. (1900)
On The Criterion That a Given System of Derivations From The Probable In The Case of a
Correlation System of Variables In Such That It Can Be Reasonably Supposed To Have Arisen From
Random Sampling.
Philosophy Magazine Series (5). 50, 157 - 172.

- [18] Tate, M. W. e Hyer, L. A. (1973)
Inaccuracy of the χ^2 Test of Goodness of Fit When Expected Frequencies Are Small.
Journal of the American Statistical Association. 68, 836 - 841.

Endereço para correspondências:

Universidade:

UFC - Centro de Ciências - D.E.M.A.

Campus do Pici - Bloco 910 - Caixa Postal: 12.183

Pici - Fortaleza - Ce

Cep: 60.455 - 760

Fone: (085) 288-9840

Residência:

Rua: Alcântara Bilhar, 677. Apto - 102, bloco J

Antônio Bezerra - Fortaleza - Ce

Cep: 60.356 - 530

Fone: (085) 286-5671

E-Mail

jalles@ufc.br