

Testes Monte Carlo Convencionais e Seqüenciais Comparação dos Poderes e dos Tempos de Execução

Ivair Silva, Renato Assunção, Marcelo Azevedo

Departamento de Estatística

Instituto de Ciências Exatas

Universidade Federal de Minas Gerais

Belo Horizonte, 02 de março, 2007

Resumo

A realização de um teste estatístico de hipóteses está condicionada à distribuição de probabilidade da estatística de teste sob a hipótese nula, pois é a partir desta distribuição que se obtém o valor-p. Em algumas situações não é possível deduzir teoricamente a distribuição da estatística de teste sob H_0 , e uma forma de conduzir o teste nestes casos é obter o valor-p via simulações Monte Carlo.

O teste Monte Carlo convencional, ou seja, que usa um número m fixo de simulações para obtenção do valor-p, pode vir a se tornar custoso computacionalmente. Uma metodologia alternativa para obter o valor-p via simulação Monte Carlo, sem ser preciso fixar o número de simulações, é o teste seqüencial. Esta abordagem faz com que o tempo de execução do teste seja esporadicamente reduzido em relação ao teste convencional.

Este trabalho tem dois objetivos principais: reduzir o tempo de execução do teste seqüencial de Monte Carlo sem que isso afete seu poder; comparar o poder do teste MC sequencial com o do MC convencional. Assim, mostrou-se que é possível estabelecer um máximo para o tempo de execução do método seqüencial, uma vez que o poder se torna constante a partir de um certo número de simulações. Mostrou-se também que é sempre melhor optar pelo método seqüencial em substituição ao método convencional, pois existe um critério de aplicação do seqüencial que garante equivalência de poder entre ambos e gera esporadicamente tempos de execução inferiores.

Para um teste seqüencial genérico, foram estabelecidas cotas para as diferenças entre os poderes deste e o do convencional. Por exemplo, para o

procedimento seqüencial que é, no mínimo, 3.3 vezes mais rápido que o convencional com $m=999$, a cota superior teórica para a perda de poder é aproximadamente 0.16, mas uma análise da curva de poder destes testes mostra que, na prática, esta perda é bem menor. Um procedimento seqüencial, que tem tempo de execução no mínimo 2 vezes inferior, apresenta cota superior teórica, para a perda de poder, em torno de 0.092, mostrando que a redução de poder ao se utilizar o método seqüencial é pequena, mesmo para situações em que seu tempo de execução é substancialmente inferior ao do convencional. Obtivemos também cotas para a diferença de poder entre estes testes Monte Carlo e o teste exato.

Palavras chaves: Teste Monte Carlo, teste seqüencial, poder, valor-p.

1 Introdução

A realização de um teste de hipóteses via simulação Monte Carlo pode se tornar expressivamente custoso computacionalmente. Uma tentativa de torná-lo mais ágil é a utilização do procedimento seqüencial. No teste seqüencial o número de estatísticas geradas sob a hipótese nula não é fixo, ao contrário do que ocorre no procedimento convencional.

Este artigo foi motivado pela idéia de se obter o valor-p do teste Monte Carlo via metodologia seqüencial, Besag & Clifford (1991). A literatura da área de genética apresenta boas aplicações do teste Monte Carlo Seqüencial, podendo-se citar Wigginton & Abecasis (2006), Bahlo et al. (2006), Browning & Floss (2006), Lindstrom et al. (2005) e Song et al. (2004).

Como trabalhos teóricos envolvendo testes MC Seqüenciais podem ser citados: o artigo que trata de um algoritmo seqüencial para os métodos de reamostragem por permutação com reposição e do método Bootstrap de reamostragem não-paramétrico com reposição, Fay & Follmann (2002); um conceito utilizado adiante, e também trabalhado em Fay & Follmann (2002), é o do valor-p como variável aleatória - uma discussão detalhada sobre este conceito pode ser obtida em Sackrowitz & Cahn (1999); em Diggle & Gratton (1984) são trabalhados métodos Monte Carlo de inferência para modelos implícitos; Fill (1998) propõe um algoritmo para amostragem perfeita por Cadeias de Markov; Besag et al. (1995) enfatiza o paradigma Bayesiano e a probabilidade condicional, mostrando sua relação com os campos aleatórios de Markov em estatística espacial.

Um aspecto ainda não esclarecido sobre a utilização do procedimento seqüencial de MC é a relação do poder deste com o do convencional no caso geral: haveria algum tipo de ônus ao se aplicar este método em substituição ao método padrão, nos casos

em que não se tem informação alguma sobre a distribuição da estatística de teste sob H_0 ? A resposta para tal questão é o principal foco deste artigo.

Este trabalho está organizado da seguinte maneira: A próxima seção contém uma síntese de idéias e notações do procedimento seqüencial de Monte Carlo proposto em Besag & Clifford (1991). Na seção 3, estão apresentados aspectos relacionados ao tempo de execução do procedimento seqüencial e, na seção subsequente, mostra-se como estipular os parâmetros do método seqüencial, de modo que exista equivalência de poder com o MC convencional. Na seção 5, são estabelecidas cotas para as diferenças de poder entre os dois métodos, e para a diferença dos poderes destes em relação ao respectivo teste exato. A seção 6 trata da estipulação de um critério para escolha sensata do número m de simulações no MC convencional. As seções 7 e 8 trazem aplicações do método seqüencial, por meio da estatística de Varredura Kulldorff & Nagarwalla (1995), principal método utilizado para detecção de conglomerados espaciais, explorando empiricamente seu poder e tempo de execução, comparado com os do MC convencional. Por último, a seção 9 faz uma discussão sobre as implicações dos resultados obtidos.

2 Teste Monte Carlo Seqüencial

Quando lidamos com a aplicação de um teste de hipóteses, o principal elemento que nos guia sobre a decisão de rejeitar ou não a hipótese nula é o valor-p. Por vezes, a distribuição de probabilidade da estatística de teste U , sob a hipótese nula, é desconhecida e a obtenção do valor-p torna-se inviável. Em alguns casos, o valor-p associado ao teste é calculado via simulação de Monte Carlo.

O procedimento de Monte Carlo gera uma amostra de tamanho $(m-1)$ da estatística de teste sob a hipótese nula. Denotemos cada estatística simulada por t_i , $i = 1, \dots, m-1$. A estatística de teste é confrontada com o histograma daquelas amostradas. Se o posto de u se encontrar entre os αm maiores postos da amostra, composta por t_1, \dots, t_{m-1}, u , rejeita-se a hipótese nula ao nível de significância $(\alpha 100)\%$, ou seja, valores grandes de u levam à rejeição da hipótese nula.

Neste procedimento convencional de Monte Carlo, o valor-p é obtido pela razão g/m , em que g é o número de valores maiores ou iguais a u , dentre os m valores t_1, \dots, t_{m-1}, u . Sob a hipótese nula, a estatística G/m , onde G é a variável aleatória associada a g , é exato, ou seja, tem exatamente a distribuição do valor-p obtido pelo teste exato sob a hipótese nula, com uma distribuição discreta uniforme em $(h/h, h/(h+1), \dots, h/m, (h-1)/m, \dots, 1/m) \subset (0, 1)$. Vamos denotar este procedimento por $MCconv(m, \alpha)$.

O parâmetro α não é necessário para especificar o procedimento, sendo útil apenas

na tomada de decisão sobre a rejeição da hipótese nula, porém, ele entra na notação porque, mais à frente, será fundamental para o estudo do poder do teste.

Por vezes, este procedimento convencional do teste de Monte Carlo, em que são geradas m estatísticas sob a hipótese nula, pode se tornar computacionalmente intensivo. Com o propósito de reduzir o tempo de execução do teste de Monte Carlo, Besag & Clifford (1991) apresentam o método seqüencial para obtenção do valor-p no teste Monte Carlo, o qual é descrito a seguir:

Denotemos por g o número de valores, em l gerados sob a hipótese nula, maiores ou iguais à estatística u . O critério de parada agora será: o procedimento será interrompido no momento em que g for igual a um valor pré-estabelecido, que denotaremos por h , ou em que l atingir $n - 1$. Em outras palavras, a geração de amostras sob H_0 será interrompida quando existirem h valores maiores ou iguais a u , ou quando já tiverem sido geradas $n - 1$ amostras. O valor-p será dado por:

$$P_f = \begin{cases} g/l & \text{se } g = h \\ (g + 1)/n & \text{se } g < h \end{cases}$$

Pode ser mostrado que este valor-P também é exato. Vamos denotar este procedimento por $MCseq(n, h, \alpha)$. Outra vez, apesar de α não ser necessário para que o procedimento seja especificado, é indispensável considerá-lo esta notação, pois terá papel fundamental mais à frente.

Vale frisar que valor h é um parâmetro arbitrário, escolhido convenientemente por quem está aplicando o método seqüencial, pois, quanto maior for h , maior também será o número de simulações para que sejam obtidos estes h valores excedendo u .

Outra notação que pode causar confusão é o número máximo de simulações do procedimento seqüencial, que no caso estamos denotando por n . Enquanto que para o Monte Carlo convencional, denotamos por m . Esta diferenciação é necessária, pois estes limites para o número de simulações, em ambos os procedimentos, não precisam ser necessariamente iguais.

A v.a. L irá denotar o número de simulações até o momento de parada. A distribuição de L com este esquema, sob a hipótese nula, é:

$$P(L \leq l) = \begin{cases} 0 & \text{se } l = 0, 1, \dots, h - 1 \\ 1 - \frac{h}{l+1} & \text{se } l = h, h + 1, \dots, n - 1 \\ 1 & \text{se } c.c. \end{cases}$$

A esperança de L sob H_0 , para este esquema:

$$E(L) = \sum_{i=1}^{n-1} P(L \geq i) = \sum_{k=h+1}^{n-1} k^{-1} \cong h + h \log \left(\frac{n-0.5}{h+0.5} \right) \quad (1)$$

Assim como foi obtida a esperança de L sob H_0 , também é possível deduzir sua variância. A medida de variabilidade, apresentada pelo número de simulações no procedimento seqüencial, também é um fator importante para avaliação das vantagens deste método.

Primeiramente iremos deduzir o segundo momento de L :

$$\begin{aligned} E(L^2) &= \sum_{i=h}^{n-1} i^2 P(L = i) = \sum_{i=h}^{n-1} i^2 [P(L \geq i) - P(L \geq i+1)] \\ &= \sum_{i=h}^{n-1} i^2 \left[\frac{h}{i} - \frac{h}{i+1} \right] = h^2 \left[\frac{h}{h} - \frac{h}{h+1} \right] + (h+1)^2 \left[\frac{h}{h+1} - \frac{h}{h+2} \right] + \\ &+ \dots + (n-1)^2 \left[\frac{h}{n-1} \right] = h^2 + \sum_{i=h+1}^{n-1} \frac{h}{i} [i^2 - (i-1)^2] \\ &= h^2 + h \sum_{i=h+1}^{n-1} \left(2 - \frac{1}{i} \right) = h^2 + 2h(n-h-1) - h \sum_{i=h+1}^{n-1} \frac{1}{i} \\ \Rightarrow E(L^2) &= h^2 + 2h(n-h-1) - E(L) + h \end{aligned} \quad (2)$$

E a variância fica:

$$\begin{aligned} Var(L) &= E(L^2) - E^2(L) = h^2 + 2h(n-h-1) - E(L) + h - E^2(L) \\ &= h(2n-h-1) - E(L)[E(L) + 1] \end{aligned} \quad (3)$$

A variância de L depende apenas de sua esperança, de n e de h , e seu valor é altamente afetado pelo valor de h .

Um resultado interessante de Besag & Clifford (1991) é a relação do erro padrão do estimador do valor-p com o valor escolhido para h , para o esquema em que o valor de L é limitado: O valor-p, denotado por p , será estimado com um erro padrão que é $c100\%$ de p , em que $c = \frac{1}{\sqrt{h}}$. Por exemplo, se $h = 25$, $c = 0.2$, ou seja, p será estimado com um erro padrão que corresponde a aproximadamente 20% de seu valor.

Vimos até aqui, que na aplicação do procedimento seqüencial, é necessário escolher arbitrariamente três parâmetros, o nível α de significância, h e n . Estas escolhas podem não ser fáceis de fazer, principalmente a de n . Vimos que o proposto originalmente era escolher n igual a m , que está associado ao MC convencional, o que parece ser uma atitude conservadora. Mas, será que é realmente necessário ir até m , ou só até m , para todo h e α ? A próxima seção se dedica a responder essa pergunta.

3 Escolha de n no MC seqüencial

Esta Seção será voltada a estabelecer uma relação entre n e os parâmetros h e α , no sentido que, ao se escolherem os dois últimos, o primeiro é estabelecido automaticamente, com a condição de que o poder permaneça constante.

No procedimento de Besag & Clifford (1991), n é escolhido de maneira arbitrária, ou seja, sua escolha não leva em conta as probabilidades dos erros tipo I e tipo II do teste. Portanto, cabe demonstrar que esta escolha pode ser otimizada, de modo a preservar as probabilidades dos erros tipo I e II. Isto é, preservando-se o nível de significância e o poder do teste.

A v.a. L pode ser entendida como uma função da variável U , associada à estatística u , de modo que a região crítica do teste de hipóteses associado também pode ser definida através de L . Quando adotamos um nível α de significância e um valor h para implementação do procedimento seqüencial, automaticamente definimos um valor crítico para rejeição de H_0 . Por exemplo, se escolhermos $h = 5$, $\alpha = 0.05$ e $n = 101$, a hipótese nula será rejeitada se l for maior ou igual a 100. Isso porque, sendo o valor-p estimado por h/l ou por $(g + 1)/n$, se $l \geq 100$, a estimativa do valor-p será com certeza menor ou igual a 0.05, implicando na rejeição de H_0 .

Vamos explorar este raciocínio considerando duas situações. Uma em que adotamos $n \geq (h/\alpha) + 1$, e outra em que adotamos $n < (h/\alpha) + 1$.

Primeiramente, vamos tratar do caso $n \geq (h/\alpha) + 1$:

A probabilidade do erro Tipo I pode ser calculada da seguinte forma:

$P(\text{Rej. } H_0 | H_0 \text{ é verd.}) = P(L \geq h/\alpha | H_0 \text{ é verd.}) = h/(h/\alpha) = \alpha$. Pois, sendo o valor-p estimado por h/l ou por $(g + 1)/n$, tal estimativa será menor ou igual a α se, e somente se, $l \geq h/\alpha$. Então, usar $n \geq (h/\alpha) + 1$, garante a probabilidade do erro tipo I em α .

Agora vamos verificar o que ocorre com o poder ainda usando $n \geq (h/\alpha) + 1$:

Vamos nomear de 'sucesso' o evento $[t_i \geq u]$, assumindo que t_i provém da distribuição sob a hipótese nula. A probabilidade de ocorrência deste evento é $P(T_i \geq u)$, que é o próprio valor-p. Assim, podemos interpretar L como sendo uma variável aleatória binomial negativa truncada em $n, n + 1, n + 2, \dots$, com probabilidade $p =$ valor-p de sucesso. A probabilidade de se ter que realizar l simulações até que existam h sucessos é:

$$P(L = l | P = p) = \begin{cases} \binom{l-1}{h-1} p^h (1-p)^{l-h} & , \quad l = h, h+1, \dots, n-2 \\ \sum_{x=n-1}^{\infty} \binom{x-1}{h-1} p^h (1-p)^{x-h} & , \quad l = n-1 \\ & , \quad c.c. \end{cases}$$

A probabilidade de rejeitar H_0 no teste seqüencial, para $n \geq \frac{h}{\alpha} + 1$, pode ser expressa por:

$$\begin{aligned} P_L(\text{Rej. } H_0 | P = p) &= P_L(L \geq (h/\alpha) | P = p) = P_L(L > (h/\alpha) - 1 | P = p) \\ &= \sum_{x=0}^{h-1} \binom{\frac{h}{\alpha} - 1}{x} p^x (1-p)^{\frac{h}{\alpha} - x - 1} \end{aligned} \quad (4)$$

pois, se L for maior ou igual a (h/α) , em $(h/\alpha) - 1$ simulações não foram observados h sucessos, ou seja, ocorreram de 0 a $(h-1)$ sucessos em uma amostra de tamanho $(h/\alpha) - 1$.

A expressão (4) não depende de n , mostrando assim que, para n a partir de $(h/\alpha) + 1$, o poder é constante.

Vamos tratar do caso $n < (h/\alpha) + 1$:

Vamos definir G como sendo a variável aleatória associada a g , uma vez que g é o n° observado de t'_i s que excedem u em $(n-1)$ simulações.

Se o procedimento for interrompido pelo fato de terem sido observados os h valores excedendo u , estima-se o valor-p por g/l . Chamemos este de critério 1 de parada. O menor valor que g/l pode assumir é $h/(n-1)$. Como $n < (h/\alpha) + 1$, $h/(n-1) > h/(h/\alpha) = \alpha$.

Ou seja, a estimativa do valor-p neste caso será sempre maior que α , e nunca rejeitaremos H_0 com o critério 1 de parada.

No caso de interrompermos as simulações por termos atingido $(n-1)$ amostras, o valor-p será estimado por $(g+1)/n$. Para rejeitarmos H_0 é necessário que $(g+1)/n \leq \alpha \Rightarrow g \leq \alpha n - 1$. Ou seja, como nunca rejeitamos H_0 pelo critério 1 de parada, rejeitamos H_0 se, e somente se, $g \leq \alpha n - 1$. A $P(G \leq \alpha n - 1) = \text{sob } H_0 = \text{int}(\alpha n)/n$, em que $\text{int}(\alpha n)$ representa o menor inteiro de (αn) . Desta forma, vemos que a probabilidade de rejeitarmos H_0 , dado que é verdadeira, é igual a α se n for múltiplo de $1/\alpha$, e é menor que α para qualquer outro n menor que $(h/\alpha) + 1$.

Quanto ao poder, temos: a probabilidade de rejeitarmos H_0 é a probabilidade de termos no máximo $\text{int}(\alpha n - 1)$ sucessos em $n - 1$ simulações:

$$\begin{aligned}
P(\text{Rej. } H_0 | P = p) &= P(G \leq \text{int}(\alpha n - 1) | P = p) = \\
&= \sum_{x=0}^{\text{int}(\alpha n - 1)} \binom{n-1}{x} p^x (1-p)^{n-1-x}
\end{aligned}$$

Do que podemos concluir que o poder não é constante para $n < (h/\alpha) + 1$, diferente do que acontece com $n \geq (h/\alpha) + 1$. Na seção 6 será feita uma discussão mais aprofundada sobre a relação de $1/\alpha$ com o poder do teste MC.

Mas, já que para $n = (h/\alpha)$ a probabilidade do erro tipo I também é α , por que não adotarmos este n ao invés de $(h/\alpha) + 1$? A resposta é que no caso $n = (h/\alpha)$, a probabilidade de rejeitar H_0 é a probabilidade de em $(h/\alpha) - 2$ simulações ocorrerem no máximo $h - 1$ sucessos, e esta probabilidade é menor ou igual à probabilidade de ocorrerem no máximo $h - 1$ sucessos em $(h/\alpha) - 1$ simulações, que é o caso $n > (h/\alpha)$.

Portanto, adotar valores para n maiores que $(h/\alpha) + 1$ é desnecessário, pois o poder será o mesmo e o tempo de execução será maior. Por exemplo, sob a hipótese nula, para $h = 5$, se utilizarmos $n = 1001$, $E(L) \approx 31$. Por outro lado, usando $n = (5/0.05) + 1 = 101$, $E(L) \approx 19$. Entretanto, o ganho mais expressivo que teríamos ao utilizarmos $n = 101$, seria nas situações em que a hipótese nula é falsa, pois o número de simulações de Monte Carlo, na pior das hipóteses, chegaria a 100, ao passo que, para valores de u extremos, se não forem interrompidas, as simulações podem chegar a 1000.

Considerando que $n < (h/\alpha)$ só preserva o nível de significância se múltiplo de $1/\alpha$, para $n = (h/\alpha)$ o poder é esporadicamente menor que para os casos em que $n > h/\alpha$ e que para $n \geq (h/\alpha) + 1$ o poder se torna constante, tem-se o critério para escolha de n :

$$n = (h/\alpha) + 1 \quad (5)$$

Foi possível estabelecer uma relação entre os parâmetros do procedimento seqüencial e seu poder, o que favoreceu-nos estabelecer a regra para escolha de n . Essa descoberta nos instiga a pensar se também não haveria uma relação entre estes parâmetros e os parâmetros do procedimento convencional, de modo que os poderes destas duas metodologias estejam relacionados. A próxima seção abordará essa idéia.

4 Equivalência entre MC convencional e seqüencial

Nesta seção iremos em busca de uma condição que garanta equivalência de poder entre os procedimentos convencional e seqüencial de Monte Carlo, por meio de uma

relação que envolva os parâmetros h , α , n e m .

Analogamente ao que foi feito na seção anterior, definiremos como 'sucesso' para o procedimento Monte Carlo convencional o evento $[t_i \geq u]$, e por $P(T_i \geq u) = \text{Valor-p}$, a probabilidade de sucesso associada a esse evento. Se g é o número de t_i 's maiores ou iguais a u , então H_0 é rejeitada se $g \leq (\alpha m - 1)$.

A probabilidade de obtermos g sucessos em uma amostra de tamanho $(m - 1)$ é:

$$P(G = g | P = p) = \binom{m-1}{g} p^g (1-p)^{m-g-1}, g = 0, 1, \dots, m-1. \quad (6)$$

Isto é, $G \text{ binomial}(m - 1, p)$. Para o procedimento Monte Carlo convencional, a probabilidade de rejeitar a hipótese nula pode ser calculada por:

$$P_G(\text{Rej. } H_0 | P = p) = P_G(G \leq \alpha m | P = p) = \sum_{y=0}^{\alpha m - 1} \binom{m-1}{y} p^y (1-p)^{m-y-1} \quad (7)$$

Com as expressões (4) e (7), podemos estabelecer uma relação entre m e os parâmetros h e α , de modo que $PMCconv(m, \alpha) = PMCseq(n, h, \alpha)$, para qualquer que seja o valor-p, independentemente de a H_0 ser verdadeira, onde $PMCconv(m, \alpha)$ e $PMCseq(n, h, \alpha)$ são os poderes dos testes convencional e seqüencial, respectivamente. Se fixarmos α e m , para m múltiplo de $1/\alpha$, as probabilidades descritas em (4) e (7) são iguais se:

$$\alpha m - 1 = h - 1 \Rightarrow h = \alpha m$$

Combinando este resultado com o da seção 3, temos que:

$$PMCconv(m, \alpha) = PMCseq\left(\frac{h}{\alpha} + 1, \alpha m, \alpha\right)$$

Para um m não múltiplo de $1/\alpha$, αm não é inteiro. Neste caso, temos que escolher um valor para h que se aproxime de αm . Independentemente de escolhermos h como menor ou maior inteiro de αm , o poder do teste seqüencial será no mínimo o do convencional. Então, a única forma de obtermos um procedimento seqüencial equivalente em poder ao convencional, é se m for múltiplo de $1/\alpha$, pois, quando isso não ocorre, e a relação $h = \alpha m$ é usada, tem-se que o teste seqüencial é mais poderoso, porém, com um tempo de execução semelhante. Na seção 6 será feita uma discussão fundamentada sobre a relação entre m e $1/\alpha$.

Por exemplo, na Tabela 1, usando $m = 1000$, um dos esquemas mais usados nas aplicações do Monte Carlo, e $h = 50$, o poder do teste seqüencial será igual ao do convencional, mas com a vantagem de estar sujeito a tempos de execução inferiores.

Tabela 1: Esquemas seqüenciais equivalentes em poder ao MC convencional, para $m = 1000$, $n = 1001$, e quartis associados à distribuição de L (L é a V.A que conta o n° de simulações no procedimento seqüencial)

α	h	1° quartil	Mediana	3° quartil	E(L)
0,01	10	12,33	19,00	39,00	55,56
0,02	20	25,67	39,00	79,00	97,74
0,03	30	39,00	59,00	119,00	134,69
0,04	40	52,33	79,00	159,00	168,24
0,05	50	65,67	99,00	199,00	199,26

A distribuição de L é fortemente assimétrica. Portanto, para avaliar e comparar tempos de execução do procedimento seqüencial é aconselhável analisar, além da média, os quartis, pois assim tem-se uma idéia melhor sobre o comportamento desses tempos.

Mostramos que existe uma regra para utilização do $MCseq(n, h, \alpha)$ que garante o mesmo poder que o $MCconv(m, \alpha)$ apresentando número de simulações substancialmente menor que m sob H_0 . Mas poderíamos preferir um esquema que, ao se admitir um pequeno ônus, aceitável, em termos de poder, nos daria um tempo de resposta ainda menor, reduzido mesmo quando estivermos sob a hipótese alternativa. Se quisermos adotar um esquema seqüencial genérico, precisamos ter controle sobre a perda de poder a que estaremos expostos ao abrirmos mão do teste convencional. A próxima seção tratará sobre como estabelecer cotas para esta perda.

5 Cotas para a diferença de poder

O $MCseq(n, h, \alpha)$ pode apresentar tempos de execução muitas vezes inferiores aos do $MCconv(m, \alpha)$, porém, também podem existir perdas de poder ao se fazer opção por ele. Esta seção estabelece cotas para estas perdas em um esquema seqüencial genérico, além de cotas em relação ao poder do teste exato.

Se quisermos calcular o poder para um dos métodos, é necessário o fornecimento da distribuição densidade de probabilidade do Valor-p sob a hipótese alternativa. Mas não é exatamente o poder que queremos avaliar, e sim a diferença entre os poderes dos procedimentos Monte Carlo Convencional e Seqüencial. Neste sentido, pensou-se em explorar as expressões (4) e (7) para desenvolver um artifício que, ao se estabelecer h e α , seja possível estipular um máximo para a diferença entre os poderes destas duas versões do teste.

Suponhamos que o Valor-p tenha uma função distribuição de densidade dada por $f_P(p)$. A probabilidade de rejeitarmos H_0 no $MCseq(n, h, \alpha)$, para todo p , é dada por:

$$\begin{aligned}
PMCseq(n, h, \alpha) &= P_L(Rej.H_0) = P(L \geq h/\alpha) = \int_0^1 P(L \geq h/\alpha|P) f_P(p) dp \\
\Rightarrow PMCseq(n, h, \alpha) &= \int_0^1 \sum_{x=0}^{h-1} \binom{\frac{h}{\alpha} - 1}{x} p^x (1-p)^{\frac{h}{\alpha} - x - 1} f_P(p) dp
\end{aligned} \tag{8}$$

Analogamente, pela expressão (7), a probabilidade de rejeitarmos H_0 no $MCconv(m, \alpha)$, para todo p , é dada por:

$$\Rightarrow PMCconv(m, alpha) = \int_0^1 \sum_{y=0}^{\alpha m - 1} \binom{m - 1}{y} p^y (1-p)^{m - y - 1} f_P(p) dp \tag{9}$$

Sob H_0 $f_P(p)$ é a densidade de uma distribuição uniforme em $(0,1)$, e as probabilidades dadas pelas expressões (8) e (9) são exatamente o nível de significância dos testes. Sob alguma hipótese alternativa fixa, qualquer que seja a função densidade que $f_P(p)$ assumir, tais probabilidades serão os poderes dos testes seqüencial e convencional, respectivamente. Então, para uma f.d.p $f_P(p)$ qualquer, a diferença entre os poderes dos procedimentos Monte Carlo convencional e seqüencial é dada por:

$$E[D(P)] = \int_0^1 \sum_{y=0}^{\alpha m - 1} \binom{m - 1}{y} p^y (1-p)^{m - y - 1} - \sum_{x=0}^{h-1} \binom{\frac{h}{\alpha} - 1}{x} p^x (1-p)^{\frac{h}{\alpha} - x - 1} f_P(p) dp \tag{10}$$

onde $D(P)$ é uma função da variável aleatória p , dada por:

$$D(P) = \left[\sum_{y=0}^{\alpha m - 1} \binom{m - 1}{y} p^y (1-p)^{m - y - 1} - \sum_{x=0}^{h-1} \binom{\frac{h}{\alpha} - 1}{x} p^x (1-p)^{\frac{h}{\alpha} - x - 1} \right] \tag{11}$$

Para a e b reais, $a \leq D(P) \leq b$ implica que $a \leq E[D(P)] \leq b$. Desta forma, podemos estipular cotas para a diferença entre os poderes dos procedimentos Monte Carlo convencional e seqüencial por meio da função $D(P)$, que pode ser facilmente calculada para um dado valor- p .

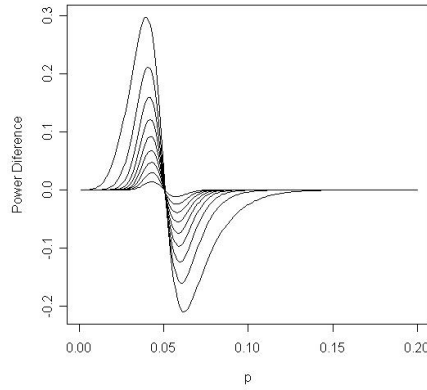
A partir da análise destas cotas, bem como dos tempos mínimo, médio e máximo de execução do procedimento seqüencial, é possível escolher os parâmetros para sua operacionalização que sejam mais convenientes e aceitáveis. Se impusermos uma grade bem fina de valores entre zero e um para P podemos aproximar estas cotas com qualquer grau de precisão (ver apêndice 1).

A Figura 1 apresenta a representação gráfica de $D(P)$ imposta a essa grade para valores de h entre 5 a 45, sendo que a curva desta Figura que apresenta o maior pico corresponde a $h = 5$, e os picos vão diminuindo à medida que aumenta h , mostrando que quanto maior h , menor a diferença de poder em relação ao MC com 999 simulações.

Veja que para $h = 5$, o mínimo de $D(P)$ se encontra em torno de $a = -0.21$ e o máximo, próximo de $b = 0.298$. $D(P)$ assume seu máximo para um valor- p próximo de 0.03967. Para que a diferença dos poderes com $h = 5$ atingisse o patamar máximo (ou seja, $b = 0.298$), seria necessário que a densidade de P fosse altamente concentrada no valor 0.03967.

Apesar de 0.298 parecer um valor alto para a cota superior da diferença, uma vez que a diferença real é dada pela ponderação de $D(P)$ por $f_P(p)$, e a variância de P é maior que zero, possivelmente tal diferença de poder é bem inferior a 0.298.

Frisemos também que existe um limiar para os valores de $D(P)$, próximo a 0.043, em que os valores de $D(P)$ passam de positivos para negativos 'rapidamente', mostrando que se a variação de P for tal, que sua densidade atinja significativamente valores em torno de 0.043, já será suficiente para trazer a esperança de $D(P)$ a valores bem inferiores a 0.298, ou melhor, a diferença de poderes é pequena para uma variância de P não muito pequena.



Para a obtenção de cotas para a diferença de poder entre o teste MC e o teste exato, basta seguir o mesmo raciocínio:

A probabilidade de rejeitar H_0 no teste exato é:

$$P(P \leq \alpha) = \int_0^\alpha f_P(p) dp = \int_0^1 1_{(0, \alpha]}(P) f_P(p) dp$$

E a diferença de poder entre os testes MC e exato é:

$$E[D^*(P)] = \int_0^1 \left[1_{(0, \alpha]}(P) - \sum_{x=0}^{\alpha m - 1} \binom{m-1}{x} p^x (1-p)^{m-x-1} \right] f_P(p) dp \quad (12)$$

onde,

$D^*(P) = \left[1_{(0,\alpha]}(P) - \sum_{x=0}^{\alpha m - 1} \binom{m-1}{x} p^x (1-p)^{m-x-1} \right]$
e $1_{(0,\alpha]}(P)$ é a função indicadora de P no intervalo $(0, \alpha]$.

Sejam a^* e b^* reais, tais que $a^* \leq D^*(P) \leq b^*$ implicando em $a^* \leq E[D^*(P)] \leq b^*$. Então, a^* e b^* definem cotas para a diferença de poder entre o procedimento MC e o teste exato, tanto para o MC convencional quanto para o seqüencial equivalente.

A função poder do procedimento MC é monotonamente decrescente, como mostra a Figura 2. Isto significa que a maior diferença, cota b^* , entre os poderes MC e exato, ocorre em $P = \alpha$, e é dada por:

$$b^* = 1 - \sum_{x=0}^{\alpha m - 1} \binom{m-1}{x} p^x (1-p)^{m-x-1} \quad (13)$$

Em um teste com 5% de significância, $\alpha = 0.05$, para procedimentos MC com $m_1 = 100$ (ou $h=5$), $m_2=1000$ (ou $h=50$) e $m_3=10000$ (ou $h=500$), as cotas superiores para as diferenças entre os poderes em relação ao teste exato são, respectivamente, 0.5550178, 0.5173695 e 0.5054916 (Veja Tabela 2 do Apêndice 1).

A Figura 2 mostra que a função poder é monótona, mas que a relação entre as 3 curvas não é. Por exemplo, para alguns valores de p a distância vertical entre as curvas de poder m_1 e m_2 é maior que a de m_2 a m_3 , mas, para outros valores, a distância m_1 a m_2 é visivelmente menor. Ou seja, não é sempre que a diferença de poder MC 99 para 999 simulações é maior que a de 999 para 9999 simulações.

O Apêndice 1 contém a ilustração das curvas da função poder para valores para $h = 5, 10, 20, 30$. Observa-se que para valores de h a partir de 20 $m \geq 400$, as curvas de poder já são quase coincidentes com a de $h = 50 (m = 1000)$.

Agora que já sabemos como obter cotas para as diferenças entre $PMCconv(m, \alpha)$ e $PMCseq(n, h, \alpha)$, ficando claro que só é possível estipular estas cotas após termos escolhido um valor para m , seria útil se tivéssemos um critério que se baseasse no poder para a escolha desse valor, e é em busca deste critério que trabalharemos na próxima seção.

6 Escolhas de m no MC convencional

O número de simulações a serem realizadas no MC convencional é algo feito de maneira arbitrária. Iremos discutir agora a possibilidade de estipular uma condição para auxiliar na escolha desse número, minimizando o grau de arbitrariedade dessa escolha.

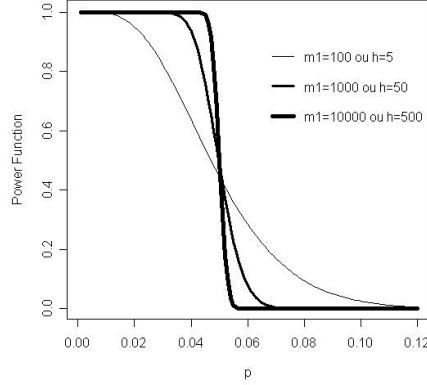


Figura 2: Função poder, $h = 5, 50, 500$

Vimos na Seção 5 que para rejeitarmos a hipótese nula no procedimento MC convencional, é necessário que sejam obtidas no máximo $(\alpha m - 1)$ t_i 's maiores ou iguais a u . Com essa relação podemos estipular um mínimo para m , partindo da idéia de que um procedimento que se utilizasse desse mínimo nos exigiria que não fosse obtido nenhum t_i maior que u , ou seja, $\alpha m - 1 = 0$, que é o mesmo que $m = 1/\alpha$. Então, para que o procedimento MC convencional tenha poder maior que zero, é necessário que m , seja no mínimo, $1/\alpha$. Por exemplo, para $\alpha = 0.05$, m deve ser no mínimo 20.

Intuitivamente pode-se pensar que, quanto maior for m , maior será o poder do MC convencional. Mas não é bem essa a relação entre o poder e m . Por exemplo:

Suponhamos que dois pesquisadores queiram realizar um teste e que para isso, ambos utilizarão simulação MC. Suponha-se que um deles escolha realizar $m - 1$ simulações. O segundo com a intenção de fazer um teste mais poderoso, escolhe realizar $m - 1 + k$ simulações. Para o primeiro rejeitar a hipótese nula, ele precisa observar no máximo $\alpha m - 1$ t_i 's maiores ou iguais a u , enquanto que o segundo poderia observar no máximo $\alpha(m + k) - 1$. Para que o poder do segundo seja maior que o do primeiro, pela expressão (9), é preciso que $\alpha(m + k) - 1$ seja maior que $\alpha m - 1$, o que equivale a $\alpha(m + k) - 1 \geq \alpha m - 1$, implicando que $k \geq \frac{1}{\alpha}$.

Concluimos, então, que um procedimento mais poderoso no MC convencional só é obtido quando aumentamos m em $1/\alpha$ unidades. Esta informação, combinada com o fato de existir um mínimo para m , nos leva a concluir que devemos sempre escolher m como múltiplo de $1/\alpha$. Este resultado justifica a afirmação feita na seção 4 sobre a escolha de m . Por exemplo, para $\alpha = 0.05$, m deve ser sempre múltiplo de 20, pois do contrário, o procedimento possivelmente terá um poder inferior a um outro

que apresente tempo de execução inferior.

Para exemplificar o que foi exposto acima, suponhamos que um indivíduo escolha $m = 1050$ e $\alpha = 0.01$, então $1/\alpha = 100$. O poder deste teste será inferior ao do teste com $m = 1000$, pois, da expressão (9):

$$\Rightarrow \int_0^1 \sum_{y=0}^9 \binom{999}{y} p^y (1-p)^{1000-y-1} f_P(p) dp \geq \int_0^1 \sum_{y=0}^9 \binom{1049}{y} p^y (1-p)^{1050-y-1} f_P(p) dp$$

Explicitando melhor, para $\alpha = 0.01$, tomando-se como ponto de partida um m igual a 1000, o poder decai até m igual a 1099, e só em m igual a 1100 o poder se torna maior que aquele que se tem com 1000. Conclui-se, então, que escolher valores para m intermediários aos múltiplos de $1/\alpha$ é sempre oneroso, uma vez que aumentam o tempo de execução do procedimento e diminuem o poder do teste em relação a um que use um m múltiplo de $1/\alpha$ imediatamente inferior.

Os resultados apresentados até agora sugerem uma opinião favorável à utilização do procedimento seqüencial de MC em substituição ao convencional. Para motivar futuros estudos empíricos que reforcem tais resultados, será descrita na próxima seção uma rápida análise empírica sobre a comparação dos poderes destas duas opções de aplicação do teste MC.

7 Análise empírica do poder

O objetivo desta seção é realizar uma comparação empírica de poder entre os procedimentos seqüencial e convencional de Monte Carlo.

Foi realizado, empiricamente, um comparativo entre os poderes das versões seqüencial e convencional do teste Monte Carlo, na aplicação da estatística Varredura puramente espacial. Para isto, utilizaram-se duas abordagens: uma baseada em dados que geramos computacionalmente, e a outra na utilização dos dados do acervo Benchmark, que podem ser obtidos no endereço www.satscan.com. Este software serve especificamente para detectar conglomerados espaciais via teste Varredura.

Ambas as abordagens mostraram que, para a estatística de varredura espacial, o poder empírico apresentado pela metodologia seqüencial foi equivalente à convencional. Todos os esquemas seqüenciais com valores para h maiores ou iguais 5 não apresentam perdas empíricas de poder maiores que 0.05, em relação ao teste convencional (ver Tabela 8 no Apêndice 2). Maiores detalhes sobre a operacionalização destas análises encontram-se no Apêndice 2.

A necessidade de avaliar o poder da metodologia seqüencial é uma conseqüência de

outro interesse que surgiu anteriormente, que é a redução do tempo de execução de um teste de hipóteses via simulações MC. Sendo assim, voltemos-nos também a avaliar empiricamente o objetivo principal, que é o tempo de execução do teste seqüencial, frente ao tempo do convencional. Ressaltemos que esta não é uma análise detalhada e altamente elaborada, e sim um chamado, motivador de trabalhos voltados especificamente à avaliação empírica deste aspecto.

8 Análise Empírica do tempo de execução

O interesse é verificar a vantagem, em termos do tempo de execução do teste Monte Carlo, em se utilizar o teste seqüencial em substituição ao convencional.

Foi feita a opção por continuar utilizando o teste Varredura espacial como referência de exemplo empírico, portanto, optou-se por trabalhar o software Satscan, que pode ser baixado de www.satscan.com.

O sistema de operação deste software possibilita o cálculo das razões mínima, média e máxima entre os tempos de execução dos procedimentos convencional e seqüencial de Monte Carlo. Isto é possível porque o software possui uma janela interativa, que imprime a mensagem de 'tarefa executada' a cada amostra gerada pelo procedimento Monte Carlo, permitindo que o usuário acompanhe, passo a passo, a execução do procedimento. Assim, é possível registrar o tempo mínimo, médio e máximo de execução do procedimento seqüencial que se teria para os dados em questão.

Foi utilizado o mapa do Estado de Minas Gerais, que é constituído por 853 municípios. Os dados foram gerados aleatoriamente, de modo a possibilitar uma análise espaço-tempo para 30 unidades de tempo.

Os resultados mostraram que a relação entre os tempos de execução dos procedimentos convencional e seqüencial é equivalente à razão obtida entre os números de simulações provenientes dos dois métodos. Por exemplo, para $h = 5$, $\alpha = 0.05$, $n = 101$ e $m = 1000$, o procedimento convencional foi executado em 82 minutos, enquanto que, para o mesmo banco de dados, o seqüencial seria operacionalizado em, no máximo, 12 minutos.

9 Discussões

O uso do teste Monte Carlo seqüencial, em substituição ao convencional, é viável e conveniente, pois, além das evidências empíricas que sugerem equivalência de poder e tempos bem inferiores de execução entre as duas versões, foi provado que, para cada procedimento convencional, existe um seqüencial equivalente em poder, porém,

nas situações em que a hipótese nula é verdadeira, o seqüencial apresenta tempos médios de execução significativamente inferiores, e nas situações em que a hipótese nula é falsa, o tempo é, no máximo, o mesmo que o apresentado pelo convencional.

Em certos casos, esquemas seqüenciais podem ser menos poderosos que o convencional, porém, com tempos de execução sempre inferiores. As análises das cotas entre as diferenças de poderes destas versões de operacionalização do teste MC mostram que, por exemplo, se usarmos $h=20$, o que resultaria em um máximo de 400 simulações, a perda de poder em relação ao convencional não seria maior que 0.122.

Tendo em vista o caráter genérico dos resultados e a facilidade de adaptar um procedimento convencional para o seqüencial, é recomendável que, para os testes de hipóteses via simulações Monte Carlo, seja usado o procedimento seqüencial em substituição àquele com número fixo de simulações. Poderíamos citar, como exemplo de substituição, o teste de varredura espacial, que poderia ser facilmente adaptado no software Satscan para obter o valor-p via metodologia MC seqüencial.

Referências

- Bahlo, M., Stankovich, J., Speed, T., Rubio, J., Burfoot, R. & Foote, S. (2006), 'Detecting genome wide haplotype sharing using snp or microsatellite haplotype data', *Human Genetics* **119**(1-2), 38–50.
- Besag, J. & Clifford, P. (1991), 'Sequential monte carlo p-value', *Department of Statistical, University of Oxford, 1 South Parks Road, Oxford OX1 3TG* .
- Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995), 'Bayesian computation and stochastic-systems', *STATISTICAL SCIENCE* **10**(1), 3–41.
- Browning & Floss, B. (2006), 'flexible ordered subset analysis for linkage mapping of complex traits', *BIOINFORMATICS* **22**(4), 512–513.
- Diggle, P. & Gratton, R. (1984), 'Monte carlo methods of inference for implicit statistical models', *Royal Statistical Society* **46**(2), 193–227.
- Fay, M. & Follmann, D. (2002), 'Designing monte carlo implementations of permutation or bootstrap hypothesis tests', *AMERICAN STATISTICIAN* **56**(1), 63–70.
- Fill, J. (1998), 'An interruptible algorithm for perfect sampling via markov chains', *ANNALS OF APPLIED PROBABILITY* **8**(1), 131–162.
- Kulldorff, M. & Nagarwalla (1995), 'A spatial scan statistic', *Biometry Branch, DCPC, National Cancer Institute. EPN 344, 6130 Executive Blvd, Bethesda MD 20892, USA and Department of Statistics, Uppsala University, 751 20 Uppsala, Sweden* .

Lindstrom, S., Wiklund, F., Ba. Jonsson, H. A., Balter, K., Brookes, A., Xu, J., Zheng, S., Isaacs, W., Adolfsson, J. & Gronberg, H. (2005), ‘Comprehensive genetic evaluation of common e-cadherin sequence variants and prostate cancer risk: strong confirmation of functional promoter snp’, *HUMAN GENETICS* **118**(3-4), 339–347.

Sackrowitz, H. & Cahn, E. S. (1999), ‘P values as random variables-expected p values’, *The American Statistician* **53**(4), 326–331.

Song, K., Weeks, D., Sobel, E. & Feingold, E. (2004), ‘Efficient simulation of p values for linkage analysis’, *GENETIC EPIDEMIOLOGY* **26**(2), 88–96.

Wigginton, J. & Abecasis, G. (2006), ‘An evaluation of the replicate pool method: Quick estimation of genome-wide linkage peak p-values’, *GENETIC EPIDEMIOLOGY* **30**(4), 320–322.

Apêndice 1

A Figura 3 mostra o comportamento da função poder para diferentes valores de m . Verifica-se que a relação entre as curvas não é monótona.

Isto sugere que podem existir casos em que a diferença entre os poderes para os testes com 999 e 99 simjlações, não seja maior que a diferença entre os testes com 9999 e 999 simulações.

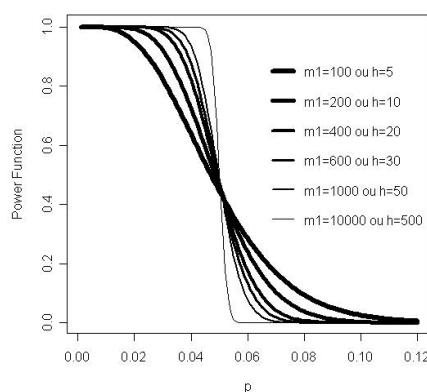


Figura 3: Curvas Função poder, $m = 100(h = 5)$, $m = 200(h = 10)$, $m = 400(h = 20)$, $m = 600(h = 30)$, $m = 1000(h = 50)$, $m = 10000(h = 500)$

A curva da função poder, para $m = 400(h = 20)$, é muito próxima à que contém $m = 1000(h = 50)$.

Tabela 2: Cotas para a diferença entre os poderes convencional e seqüencial de Monte Carlo com $\alpha = 0.05$, $m = 1000$ e $n = (h/\alpha) + 1$

h	b	PMC_{conv}	PMC_{seq}	TC/TS máximo	TC/TS médio	TC/TS mínimo
5	0,298	0,941	0,643	200	50	9,89
6	0,276	0,935	0,660	167	42	8,26
7	0,257	0,930	0,673	143	36	7,09
8	0,240	0,925	0,685	125	31	6,20
9	0,225	0,920	0,695	11	28	5,52
10	0,212	0,916	0,704	100	25	4,97
11	0,200	0,912	0,713	91	23	4,52
12	0,188	0,908	0,720	83	21	4,15
13	0,178	0,905	0,727	77	19	3,83
14	0,168	0,902	0,733	71	18	3,56
15	0,159	0,898	0,739	67	17	3,32
16	0,151	0,895	0,744	62	16	3,11
17	0,143	0,892	0,749	59	15	2,93
18	0,136	0,889	0,754	56	14	2,77
19	0,128	0,887	0,758	53	13	2,62
20	0,122	0,884	0,762	50	13	2,49
21	0,115	0,881	0,766	48	12	2,37
22	0,109	0,879	0,770	45	11	2,27
23	0,103	0,876	0,773	43	11	2,17
24	0,098	0,874	0,777	42	10	2,08
25	0,092	0,872	0,780	40	10	1,99
26	0,087	0,870	0,783	38	10	1,92
27	0,082	0,868	0,786	37	9	1,85
28	0,077	0,866	0,789	36	9	1,78
29	0,072	0,864	0,791	34	9	1,72
30	0,068	0,862	0,794	33	8	1,66
31	0,063	0,860	0,796	32	8	1,61
32	0,059	0,858	0,799	31	8	1,56
33	0,055	0,856	0,801	30	8	1,51
34	0,051	0,854	0,803	29	7	1,47
35	0,047	0,853	0,805	29	7	1,43
36	0,044	0,851	0,808	28	7	1,39
37	0,040	0,849	0,809	27	7	1,35
38	0,036	0,848	0,811	26	7	1,31
39	0,033	0,846	0,813	26	6	1,28
40	0,030	0,845	0,815	25	6	1,25
41	0,026	0,843	0,817	24	6	1,22
42	0,023	0,842	0,819	24	6	1,19
43	0,020	0,840	0,820	23	6	1,16
44	0,017	0,839	0,822	23	6	1,13
45	0,014	0,837	0,823	22	6	1,11
46	0,011	0,836	0,825	22	5	1,08
47	0,008	0,835	0,826	21	5	1,06
48	0,005	0,833	0,828	21	5	1,04
49	0,003	0,832	0,830	20	5	1,02
50	0,000	-	-	20	5	1,00

Legenda da Tabela 2:

TC/TS : Tempo de execução do procedimento convencional dividido pelo tempo de execução do seqüencial;

PMC_{conv} : Poder do teste MC convencional para o caso em que a cota b ocorre;

PMC_{seq} : Poder do teste MC seqüencial para o caso em que a cota b ocorre.

A informação fornecida pela Tabela 3 é similar à da Tabela 2, com a diferença de que, desta vez, estão estabelecidas as cotas para a diferença de poder entre os procedimentos com m variando de 100 ($h = 5$) a 10000 ($h=500$), realizando uma comparação dois a dois entre todos eles.

O sentido da comparação ilustrado na Tabela 3 é os poderes dos procedimentos da linha menos os da coluna, nesta ordem.

Tabela 3: Cotas para as diferenças de poderes entre procedimentos MC

h	m	5	10	15	20	25	30	35	40	45	50	250	500
		100	200	300	400	500	600	700	800	900	1000	5000	10000
5	100	0											
10	200	0.1045	0										
15	300	0.1601	0.0582	0									
20	400	0.1969	0.0978	0.0402	0								
25	500	0.2238	0.1273	0.0707	0.0307	0							
30	600	0.2447	0.1507	0.0949	0.0553	0.0248	0						
35	700	0.2615	0.1698	0.1150	0.0758	0.0455	0.02080	0					
40	800	0.2756	0.1859	0.1319	0.0932	0.0631	0.0386	0.0178	0				
45	900	0.2876	0.1997	0.1466	0.1084	0.0785	0.0541	0.0335	0.0156	0			
50	1000	0.2980	0.2118	0.1595	0.1217	0.0921	0.0679	0.0473	0.0296	0.0139	0		
250	5000	0.4195	0.3596	0.3221	0.2942	0.2716	0.2526	0.2361	0.2215	0.2084	0.1965	0	
500	10000	0.4537	0.4034	0.3722	0.3488	0.3298	0.3137	0.2997	0.2872	0.2760	0.2657	0.0858	0
-	Exato	0.5550	0.5389	0.5317	0.5275	0.5246	0.5224	0.5208	0.5194	0.5183	0.5174	0.5078	0.5055

Apêndice 2

A) Casos Gerados

Para explorar o poder empírico da estatística de Varredura Espacial Seqüencial, o primeiro passo foi modificar o procedimento Monte Carlo adaptando-o ao esquema seqüencial, apresentado na seção 3, em que o valor de h é fixado e o procedimento é continuado até que existam h estatísticas maiores ou iguais a u , ou até que l atinja 999.

Para as simulações, escolheu-se um dos mapas (coordenadas longitudinais e latitudinais) oferecidos pelo acervo de ajuda do software Satscan. Este mapa representa uma região composta por 32 áreas.

Foram realizados vários cenários, que se diferenciavam quanto ao valor de e e da relação entre as probabilidades p e q , as probabilidades de ocorrência de um evento dentro do cluster e fora do cluster, respectivamente.

O algoritmo de simulação do teste de varredura espacial seqüencial foi implementado no software R 2.2.1. Tal algoritmo se baseou nas seguintes etapas:

1- Tomemos as notações p e q que definem as probabilidades de ocorrência do evento dentro do cluster e fora do cluster, respectivamente. Após definir um total de casos e um total populacional, foram geradas 1000 amostras aleatórias e distribuídas no mapa, induzindo-se assim, para cada amostra, um cluster de 11 áreas segundo uma distribuição multinomial, obedecendo a uma razão p/q . Veja que para $p/q = 1$, estamos gerando uma amostra sob a hipótese nula. Quanto maior for p/q , maior ênfase será dada à hipótese alternativa;

2- Aplicou-se, para cada amostra gerada, o procedimento de Varredura Espacial obedecendo a um cenário específico. Ou seja, obteve-se para cada amostra um valor- p , resultando em 1000 valores- p ;

3- Para cada aplicação, contabilizou-se o percentual de vezes que os valores-p foram inferiores a 0.05 (pois assim estaremos adotando um nível de significância de 5%), o que pode ser visto como uma estimativa do poder.

A Tabela 4 mostra os resultados obtidos com as simulações para uma população total de 6400 indivíduos e total de casos igual a 64. São apresentadas as estimativas do poder e da média de L para cada cenário. Veja que cada linha desta Tabela é o resultado de 1000 amostras geradas, em que para cada uma destas amostras foi aplicada a estatística de varredura.

Ao que parece, o teste de Varredura seqüencial se mostrou compatível ao de Varredura convencional em termos de poder, porém, menos dispendioso computacionalmente.

O próximo passo é avaliar se estes resultados também se verificam em exemplos mais realistas, onde as situações são bem menos comportadas que os casos simulados acima, o que será tratado no próximo tópico.

B) Benchmark - (mixed01.sim, 245 áreas, 600 casos)

Neste tópico o objetivo foi comparar o poder entre os procedimentos seqüencial e convencional de MC aplicados à estatística de Varreduras com interação espaço-tempo. O que difere esta análise da do tópico A, além da introdução do tempo como fator de análise, é que naquele caso os dados foram gerados pelos autores deste artigo, enquanto que neste tópico foi utilizado o banco de dados Benchmark, que pode ser obtido no endereço www.satscan.com. Este banco foi criado e disponibilizado livremente na internet para ser usado por pesquisadores que desenvolvem ferramentas de detecção de conglomerados espaciais, justamente para que estes possam contar com uma base comum de dados para pesquisa sobre este assunto.

A base de dados, composta por amostras geradas sob a hipótese nula, possui 99.999 configurações distintas para um total de 600 casos, que são alocados a uma região que possui 245 áreas. Novamente usou-se o software R 2.2.1 para implementação do algoritmo de simulação. Tomando apenas os mil primeiros cenários deste banco, o teste seqüencial, para $h = 5$ e $\alpha = 0.05$, levou à rejeição da hipótese nula em 4% das vezes. Escolhendo-se os dez mil primeiros cenários, o teste seqüencial fez com que a hipótese nula fosse rejeitada 5.3% das vezes. Isso mostra que o procedimento seqüencial preservou o nível de significância de 5% para estes cenários.

Também são oferecidos pelo referido site, vários arquivos com amostras geradas sob a hipótese alternativa.

Para estudar o poder do teste de varredura seqüencial, foi usado o arquivo mi-

xed01.sim do banco Benchmark, em que para cada amostra foi induzido apenas um conglomerado. Este arquivo contém dez mil configurações de casos. Aplicando-se o procedimento seqüencial, para as mil primeiras configurações, a hipótese nula foi rejeitada 87.4% das vezes, e aplicando-se o teste de varredura espacial convencional, rejeitou-se a hipótese nula para 88.4% das configurações. Uma diferença de 0,01. Novamente temos uma evidência sobre a compatibilidade, em termos de poder, dos testes seqüencial e convencional.

Tabela 4: Estimativas de poder para o teste de Varredura puramente espacial seqüencial - total de casos corresponde a 1% da população total

Cenário	h	p/q	P(Rej. H ₀ —p/q)	Média de l
1	1	2,00	0,350	122,476
2	2	1,60	0,199	72,498
3	2	1,70	0,223	75,385
4	2	1,80	0,274	103,932
5	2	1,90	0,324	131,002
6	2	2,00	0,382	170,797
7	5	1,00	0,056	33,086
8	5	1,10	0,056	34,218
9	5	1,20	0,072	41,498
10	5	1,30	0,074	49,257
11	5	1,40	0,097	62,229
12	5	1,50	0,124	71,380
13	5	1,60	0,191	114,847
14	5	1,70	0,237	137,855
15	5	1,80	0,272	168,501
16	5	1,90	0,318	193,598
17	5	2,00	0,413	256,014
18	10	1,00	0,045	54,617
19	10	1,10	0,042	54,535
20	10	1,20	0,068	63,837
21	10	1,30	0,074	77,225
22	10	1,40	0,093	86,911
23	10	1,50	0,132	119,956
24	10	1,60	0,192	166,854
25	10	1,70	0,247	214,788
26	10	1,80	0,267	225,695
27	10	1,90	0,354	287,672
28	10	2,00	0,413	345,843
29	15	1,00	0,047	74,972
30	15	1,10	0,055	82,839
31	15	1,20	0,066	91,892
32	15	1,30	0,087	106,344
33	15	1,40	0,107	128,185
34	15	1,50	0,135	146,653
35	15	1,60	0,176	196,841
36	15	1,70	0,228	227,956
37	15	1,80	0,297	289,917
38	15	1,90	0,351	334,003
39	15	2,00	0,393	384,843
40	20	1,00	0,057	104,434
41	20	1,10	0,066	107,333
42	20	1,20	0,071	123,286
43	20	1,30	0,082	129,634
44	20	1,40	0,101	146,265
45	20	1,50	0,160	201,891
46	20	1,60	0,189	240,136
47	20	1,70	0,262	289,542
48	20	1,80	0,280	313,896
49	20	1,90	0,369	394,967
50	20	2,00	0,393	424,636
51	MC m=1000	1,00	0,055	999,000
52	MC m=1000	1,10	0,049	999,000
53	MC m=1000	1,20	0,060	999,000
54	MC m=1000	1,30	0,069	999,000
55	MC m=1000	1,40	0,113	999,000
56	MC m=1000	1,50	0,140	999,000
57	MC m=1000	1,60	0,174	999,000
58	MC m=1000	1,70	0,231	999,000
59	MC m=1000	1,80	0,295	999,000
60	MC m=1000	1,90	0,352	999,000
61	MC m=1000	2,00	0,424	999,000