INFERÊNCIA EM ECOLOGIA: COMENTÁRIOS E UM EXEMPLO

HUMBER AGRELLI ANDRADE[1] & PAUL GERHARD KINAS[2]
[1]Univ. Fed. Santa Catarina (UFSC); Depto. Informática e Estatística (INE); Depto. Matemática; Caixa Postal 476; CEP 88010970; Florianópolis-SC; humber@inf.ufsc.br
[2]Univ. Fed. do Rio Grande (FURG); Caixa Postal 474; CEP 96201900; Rio Grande – RS; dmtkinas@furg.br

RESUMO
A estimativa de intervalos de confiança e a realização de testes de hipóteses são componentes da inferência estatística fundamentais em estudos na área de ecologia quantitativa. Os procedimentos de inferência são freqüentemente utilizados por pesquisadores nas avaliações de questões e afirmações de cunho científico. Pelo menos três procedimentos podem ser utilizados para a realização de inferências. Neste trabalho avaliamos estes procedimentos com a análise de um exemplo de ecologia. Concluímos que a abordagem bayesiana é vantajosa, pois propicia que o pesquisador obtenha respostas diretas para as questões relevantes de um estudo de ecologia.

PALAVRAS-CHAVE: intervalo de confiança, teste de hipótese, análise Bayesiana

ABSTRACT
Inference in ecology: comments and an example
Statistical confidence intervals and hypotheses tests are important components of ecological studies. Inference tools are often used by researchers when they assess scientific questions and statements. At least three procedures could be used in inference. In this paper we evaluated those procedures when analyzing an example of ecology. We show that the Bayesian approach is in advantage because it allows the researchers to easily answer the relevant questions in ecological studies.

KEYWORDS: confidence interval, hypothesis test, Bayesian analysis

INFERÊNCIA

Em diversas situações no decorrer do estudo de um problema ecológico tem-se interesse em tirar conclusões sobre um "todo" (população) a partir do estudo de uma "parte" (amostra) ou, fazer previsões para o futuro com base na experiência acumulada no passado. Por exemplo, poderíamos estar interessados em avaliar o peso médio de todos os organismos de uma dada espécie, a partir de informações sobre os pesos de alguns destes organismos, ou prever a resposta dessa espécie a determinado impacto ambiental. Esse processo de estudar um caso específico e a subseqüente tentativa de se tirar conclusões sobre o caso mais genérico é denominado "indução". O processo inverso, em que se parte de um conhecimento genérico para tirar uma conclusão sobre uma especificidade, é denominado "dedução". O principal campo de aplicação do procedimento dedutivo é o da matemática pura, que se ocupa com a elaboração de conseqüências a partir de regras pré-estabelecidas. Por exemplo, se nós sabemos que 30% dos inivíduos de uma dada espécie apresentam determinada mutação genética, o processo dedutivo, com o uso de uma distribuição de probabilidade binomial, nos permite dizer que em uma amostra de 10 indivíduos a probabilidade de que 5 ou mais apresentem a mutação é 4,7%. No entanto, na maioria dos estudos ecológicos não temos de fato acesso a população toda e temos então que lidar com o desafio de fazer generalizações a partir de uma amostra. No exemplo acima, isto equivale a responder a mesma pergunta sem, no entanto, conhecer a proporção de mutações na população. Isto implica que na prática científica temos que utilizar a lógica indutiva (raciocinar da parte para o todo) ao invés da lógica dedutiva (raciocinar do todo para a parte).

Quando utilizamos algumas técnicas e procedimentos para se fazer uma indução, dizemos que estamos fazendo uma inferência estatística. O papel da inferência estatística na ciência é atualmente crucial (Salsburg, 2002). Na época contemporânea praticamente nenhuma descoberta, hipótese ou pretenso avanço é considerado cientificamente válido se não estiver referendado por inferências estatísticas, que podem ser realizadas a partir de diferentes abordagens. Este trabalho apresenta uma comparação entre essas abordagens, ilustrando-as com um exemplo prático.

A ABORDAGEM ORTODOXA

A inferência estatística amplamente utilizada pelos pesquisadores da área de ecologia, é por vezes denominada de "freqüentista", "clássica", "tradicional", ou "ortodoxa" (Jaynes, 2003). Esta é a abordagem dominante nos livros textos utilizados na grande maioria dos cursos de graduação, incluindo aqueles vinculados às ciências naturais como biologia e ecologia. Enquadra-se no contexto ortodoxo a construção de “intervalos de confiança” e a realização de “testes de hipóteses”. A construção destes dois elementos é fundamental na ciência contemporânea e está baseada na teoria de probabilidades.

A adoção do arcabouço ortodoxo não pode ser feita de maneira universal no estudo de problemas ecológicos. Há muitas vezes violações de suposições importantes. Adaptações de conceitos acabam sendo feitas para adequação dos métodos ortodoxos na solução de questionamentos práticos. Há freqüentemente uma interpretação equivocada dos resultados por parte dos pesquisadores das áreas de ciências naturais, atribuindo aos resultados ortodoxos o significado que eles gostariam que tivessem, e não o significado que efetivamente têm, e que muitas vezes podem ser de pouca utilidade para o pesquisador.

Quatro pontos fundamentais resultam em uma frustrada e desgastada relação de pesquisadores das ciências naturais com a abordagem ortodoxa de análise estatística:

Probabilidade – Na abordagem ortodoxa a probabilidade é um atributo de um fenômeno. Por exemplo, tome-se um dado honesto de seis faces. Antes mesmo de fazermos um lançamento já conhecemos a probabilidade de que a face com o número quatro caia voltada para cima. Essa probabilidade é 1/6, e na abordagem ortodoxa ela é intrínseca ao dado. Porém há fenômenos na natureza que não têm uma probabilidade intrínseca intuitiva desse tipo. Por exemplo, qual é a probabilidade de que determinada espécie de alga marinha encrustrante domine a superfície de um costão rochoso? A probabilidade é intríseca ao fenômeno? Ela está naturalmente implícita no conjunto “espécie de alga” e “costão rochoso”? Certamente não. Portanto, no contexto ortodoxo não faria sentido falar-se em probabilidades neste caso. Apesar disso, percebemos que a pergunta é legítima e muito natural na visão de um ecólogo.

Replicabilidade do experimento – Essa questão está no âmago da palavra “freqüentista”, outra das designações da abordagem ortodoxa. A probabilidade, quando entendida como um atributo, poderia ser verificada a partir da replicação do fenômeno. Aqui podemos substituir a palavra fenômeno por experimento para simplificação, sem o ônus de se perder o sentido. Por exemplo, se o experimento é o lançamento do dado, então na abordagem ortodoxa assume-se que esse experimento pode ser repetido com perfeição inúmeras vezes. A freqüência relativa com que aparece a face com o número quatro voltada para cima deveria então confirmar a probabilidade intrínseca de 1/6, como um atributo do dado. No entanto, esse processo de contagem da freqüência de ocorrências em experimentos idênticos não é aplicável em problemas ecológicos para os quais a possibilidade de replicação está ausente. É impossível replicar 1000 vezes e em condições idênticas a ocorrência da espécie de alga no costão rochoso para obter a freqüência relativa em que ela foi a espécie dominante.

Intervalos de confiança – Os intervalos de confiança calculados com a abordagem ortodoxa estão entre os recursos de análise estatística mais comuns em artigos científicos. Infelizmente são também interpretados de maneira equivocada na maioria esmagadora dos estudos relacionados às ciências naturais. Vejamos isto através de um exemplo. Suponha que um pesquisador estimou que o intervalo de 95% de confiança para a proporção de fêmeas em uma população animal seja marcado pelos limites de 0,3 e 0,4.

Poderíamos agora listar muitas interpretações equivocadas usualmente encontradas de maneira explícita ou implícita em artigos científicos. No entanto, vejamos somente quatro delas a título de exemplo. O intervalo de confiança (IC) disposto acima não está dizendo que a proporção de fêmeas da população é variável e tem 95% de probabilidade de “cair” entre 0,3 e 0,4. Este IC também não significa que o valor mais provável para a proporção de fêmeas da população é de 0,35, o valor intermediário entre os limites inferior e superior do intervalo. Este IC também não quer dizer que se fôssemos retirar outra amostra da população, a proporção de fêmeas teria 95% de probabilidade de estar entre 0,3 e 0,4. Por fim, este IC também não quer dizer que a proporção populacional de fêmeas da população é fixo, mas que 95% de todas as possibilidades estão concentradas entre 0,3 e 0,4.

O intervalo de confiança significa que se fôssemos capazes de repetir o mesmo experimento (a retirada de uma amostra aleatória de mesmo tamanho da mesma população) um grande número de vezes, obteríamos um grande número de diferentes intervalos de confiança, e que 95% deles conteriam o valor real da proporção de fêmeas da população. Usualmente segue-se então com a interpretação de que o único intervalo de confiança efetivamente calculado tem uma probabilidade de 95% de ser um dos que contém o verdadeiro valor da proporção populacional de fêmeas. Note que toda a afirmação probabilística é feita sobre o intervalo de confiança, e pouco se pode dizer sobre a proporção de fêmeas da população, que é o que realmente nos interessa.

Teste de hipótese – Usualmente em ecologia estamos interessados em confrontar uma ou mais proposições que configuram as nossas hipóteses. A intenção é, por exemplo, avaliar os dados e concluir que a hipótese A é mais plausível do que B, que por sua vez é mais plausível do que C. Também é desejável quantificar, dado toda a informação disponível, qual é a plausibilidade das diferentes hipóteses. Esse tipo de resultado não pode ser construído com o uso de testes de hipóteses ortodoxos. Nestes testes se parte do princípio de que há uma hipótese de referência supostamente correta (a hipótese nula – H_0), representada por uma igualdade. Segue-se com o cálculo da probabilidade de se retirar de uma população cuja verdade é H_0, uma amostra com características mais extremas que a amostra observada. Rejeita-se H_0 e considera-se que há evidências a favor de uma hipótese alternativa, somente se essa probabilidade é “excepcionalmente baixa” (usualmente inferior a 5%). No entanto, não se pode dizer que H_0 é mais plausível que a hipótese alternativa, mesmo quando H_0 não é rejeitada! Pois, note que a probabilidade faz afirmações sobre os dados e não sobre as hipóteses. Esse é um exemplo das constatações por vezes frustrantes acerca dos testes de hipóteses ortodoxos. Em muitos estudos sobre ecologia, o pesquisador ao fazer um teste de hipóteses, percebe que o teste é incapaz de fornecer a resposta que procurava.

A ABORDAGEM BAYESIANA

A designação dada a este procedimento é motivada pelo trabalho de Bayes (1763). Na abordagem bayesiana a probabilidade de qualquer evento, fenômeno, proposição ou hipóteses é definida como a quantificação do grau de plausibilidade. Assim tornam-se perfeitamente legítimas perguntas tais como “qual é a probabilidade de que a espécie de alga marinha A seja dominante em um determinado costão rochoso?”, que são inconcebíveis na abordagem ortodoxa. A probabilidade na abordagem bayesiana é, portanto, a “probabilidade” que pesquisadores de áreas de ciências naturais procuram, sem deixar de incluir os casos abordados de maneira frequentista como casos particulares. A conceituação Bayesiana bem mais ampla de probabilidades se encaixa mais naturalmente aos estudos ecológicos, onde a replicabilidade de experimentos raramente é factível.

Na inferência bayesiana, é construída uma distribuição de probabilidade para representar as plausibilidades relativas dos diversos valores do parâmetro de interesse, como por exemplo, a proporção de fêmeas na população. Assim o cálculo de intervalos de confiança pode ser feito a partir da exploração das características desta distribuição. Se o intervalo de confiança bayesiano (95%) para a proporção é limitado por 0,3 e 0,4, isto quer dizer exatamente que a probabilidade de a proporção de fêmeas ser um valor entre 0,3 e 0,4 é 95%.

Testes de hipóteses na abordagem bayesiana podem ser realizados de forma a fornecer respostas diretas aos questionamentos dos cientistas. Simplesmente, à luz de todas as informações disponíveis, calculam-se as probabilidades (plausibilidades) de proposições alternativas. Por exemplo, poderíamos simplesmente calcular a probabilidade de que o peso médio de um organismo seja maior que 5 kg, e compará-la diretamente com a probabilidade de que o peso seja inferior a 3 kg, para identificar qual das duas hipóteses (proposições) é mais plausível. Este tipo de teste de hipóteses está em consonância com as necessidades em estudos de ecologia e responde diretamente aos questionamentos dos cientistas.

Apesar de parecer ser a alternativa mais adequada em estudos de ecologia, a abordagem bayesiana continua sendo menos utilizada que a ortodoxa. Como citado anteriormente, um dos prováveis motivos é histórico, uma vez que ela não é usualmente ensinada em cursos de graduação. No entanto, isto ocorre não por que ela seja uma “novidade”, pois os alicerces da abordagem bayesiana são tão antigos quanto os da abordagem ortodoxa. Os estatísticos ortodoxos não aceitam a definição de probabilidade como uma medida de plausibilidade, por temerem que isto tornaria subjetiva a investigação científica. No entanto, há contra-argumentações contundentes a esta crítica, que garantem respeitabilidade científica à abordagem bayesiana. Em contrapartida, as características atrativas que ela disponibiliza ao cientista têm sido o motivo da sua rápida expansão nas mais diferentes áreas de ciências (Kinas & Andrade, 2007). No próximo item, à medida que descrevemos os componentes da formulação de Bayes, incluindo a distribuição *a priori*, voltaremos a apontar essas críticas e seus contra-argumentos com maiores detalhamentos.

ELEMENTOS DA ABORDAGEM BAYESIANA

A abordagem bayesiana deve seu nome à importância do Teorema de Bayes. A aplicação deste teorema inclui o produto de uma função de verossimilhança $p(dados \mid \theta)$, com uma distribuição *a priori* $p(\theta)$ para se obter uma distribuição *posterior* para o objeto de interesse (θ), que pode ser um escalar, um vetor, ou mesmo elementos de maior dimensão. Na formulação de Bayes,

$$p(\theta \mid dados) = \frac{p(\theta) \cdot p(dados \mid \theta)}{\int p(\theta) \cdot p(dados \mid \theta)\, \mathrm{d}\theta} \quad (1)$$

a integral do denominador é uma constante de normatização. Este termo, por vezes difícil de ser avaliado analítica ou numericamente, garante que a distribuição *posterior* integre a 1, mas não é essencial para o delineamento da sua característica mais importante que é a forma . Assim freqüentemente a avaliação da proporcionalidade:

$$p(\theta \mid dados) \propto p(\theta) \cdot p(dados \mid \theta) \quad (2)$$

é suficiente para obter a solução desejada. Segue uma descrição dos três componentes fundamentais desta solução.

Verossimilhança $p(dados \mid \theta)$

Neste componente estão todas as informações relevantes sobre θ contidas nos estudos observacionais e/ou experimentais do trabalho ($dados$). A obtenção da verossimilhança começa com a proposição de um modelo para o cálculo da probabilidade de que os dados tenham sido gerados por um θ em particular. Este valor da densidade de probabilidade é então interpretado às avessas como uma medida do quão verossímel é o valor particular de θ, à luz dos dados efetivamente disponíveis. Por exemplo, suponhamos que coletamos uma amostra de quarenta organismos cuja média de peso é $\bar{x} = 5{,}5$ kg. Poderíamos estabelecer que o modelo de probabilidade para a distribuição das médias amostrais é normal com média μ e variância σ^2 / n. O vetor $\theta = \{\mu, \sigma^2\}$ é o elemento de interesse. Para cada θ em particular, por exemplo $\theta_1 = \{\mu = 6, \sigma^2 = 2\}$, podemos calcular uma densidade de probabilidade, que neste caso é $p(dados \mid \theta_1) \approx 0{,}146$, que é, de certa forma, uma medida do grau de plausibilidade (verossimilhança) de $\theta_1 = \{\mu = 6, \sigma^2 = 2\}$ frente aos dados $n = 40$ e $\bar{x} = 5{,}5$ kg. Para $\theta_2 = \{\mu = 5, \sigma^2 = 3\}$ temos que $p(dados \mid \theta_2) \approx 0{,}275$, e portanto, diríamos que θ_2 é mais verossímel que θ_1. Encontrar o valor $\hat{\theta}$ que produz a máxima verossimilhança é um critério para obter um estimador para θ. Tais cálculos podem ser facilmente feitos com o uso de programas computacionais estatísticos (e.g. R, S-Plus, SAS, SPSS, Statistica) ou mesmo de algumas planilhas eletrônicas. Neste trabalho, todos os cálculos e gráficos foram produzidos com o programa R (R

Development Core Team, 2007), que é de livre distribuição.

É importante mencionar que o cálculo da verossimilhança, é também um elemento chave na estatística ortodoxa. Portanto não há neste ponto uma divergência entre as duas abordagens. Note também que foi preciso escolher um modelo de probabilidade para descrever os dados, um modelo normal naquele exemplo, o que não deixa de ser uma atitude parcialmente subjetiva no seio da abordagem ortodoxa.

Distribuição *a priori* $p(\theta)$

Na abordagem bayesiana uma distribuição de probabilidade para θ deve ser especificada a *priori*, levando-se em conta toda a informação que não está contida nos $dados$ obtidos na amostragem realizada no estudo, mas que tem relevância para o fenômeno em estudo. Estas informações podem provir de trabalhos observacionais e experimentais pretéritos, mas informações científicas que não tem caráter estatístico podem também ser utilizadas. O ponto é que todas as informações relevantes sobre a questão devem ser formalmente incorporadas. Só assim poderá ser realizada uma análise coerente. A desconsideração consciente de informações úteis, esta sim fere o princípio na isenção e objetividade científica.

As distribuições *a priori* estão no centro da polêmica entre bayesianos e ortodoxos. Ortodoxos alegam que, ao analisar os fragmentos de informações disponíveis para a construção da *priori*, a interpretação que um pesquisador dá a estas informações pode ter caráter subjetivo, o que seria condenável em uma análise científica. No entanto, há uma série de contra-argumentações à reticência ortodoxa ao uso da distribuição *a priori*:

i - Se todas as informações disponíveis são analisadas por diferentes pesquisadores, e todos eles fazem uso da mesma base de raciocínio lógico, as distribuições *a priori* especificadas resultam similares (Jaynes, 2003). Portanto boas distribuições *a priori* provém de análises essencialmente objetivas e não subjetivas como alegam os ortodoxos.

ii – Não há substituto para um bom banco de dados; nisso a abordagem ortodoxa e bayesiana concordam. Porém, alguns pesquisadores são partidários de que não se deve temer o uso de opiniões de especialistas enquanto não se dispõe de informações mais específicas, e que talvez aí esteja a força da análise bayesiana, que permite integrar formalmente na mesma estrutura de análise todo o conhecimento existente sobre o problema (Berger, 1985). Caso contrário parte da informação estaria sendo deliberadamente ignorada, o que não é adequado no raciocínio lógico científico principalmente nas situações em que as análises requerem uma tomada de decisão imediata. Portanto, nestes casos a incorporação na análise de "todo" o conhecimento disponível é desejável.

iii - Diferentes distribuições *a priori* podem ser especificadas, de tal forma que sempre se pode fazer uma análise do seu impacto sobre as conclusões. Em particular pode-se especificar uma *a priori* difusa, que carrega pouca ou nenhuma informação sobre θ, de tal forma que se deixa o dado "falar por si mesmo" a partir da função de verossimilhança. Assim, a distribuição *a priori* viria a ser só um elemento estrutural e não teria uma influência importante nos resultados finais.

iv - A escolha do modelo para a estruturação da verossimilhança também é subjetiva. A verossimilhança faz parte de ambas as abordagens, bayesiana e ortodoxa. Portanto não é legítimo alegar que uma se distingue da outra quanto à subjetividade envolvida.

Distribuição *posterior* $p(\theta \mid dados)$

Esta distribuição representa o compromisso entre a distribuição *a priori* (toda a informação que havia antes do estudo) e a verossimilhança (informação nova obtida com o estudo). Se os dados obtidos com o estudo são muito informativos acerca de θ, a *posterior* será dominada pela função de verossimilhança. Caso contrário a distribuição *a priori* pode influenciar substancialmente a *posterior*. Na verdade todo este processo dinâmico em que se procura um balanço entre os dados coletados e as demais informações disponíveis, propicia uma formalização quantitativa do raciocínio indutivo que todo cientista já faz intuitivamente e em termos qualitativos.

Na construção da distribuição *posterior* a integração do denominador da equação 1 pode impor algumas dificuldades técnicas. Muitas vezes, principalmente em problemas que envolvem a inferência de muitos parâmetros, não é possível uma solução analítica e abordagens numéricas, tais como simulações de Monte Carlo, são necessárias para a exploração da distribuição *posterior*. Em muitas situações de relevância prática, o uso dessas técnicas não excede em dificuldade ao uso do método de *bootstrap*, por exemplo, que já é bastante popular na abordagem ortodoxa. No entanto, em alguns casos a solução bayesiana é facilitada pelo uso de famílias conjugadas de modelos para a distribuição *a priori* e a verossimilhança, o que implica que a *posterior* assume uma forma conhecida, similar à da distribuição a *priori* e que os cálculos envolvidos se reduzem a simples atualização dos parâmetros dessa distribuição a partir de fórmulas matemáticas explícitas. Há uma série de possibilidades de conjugação, e a opção por uma ou outra alternativa depende do reconhecimento da solução que é mais adequada para o problema em análise. A identificação da solução adequada exige que o pesquisador tenha conhecimento dos fundamentos básicos de probabilidade e alguma familiaridade com as propriedades e estruturas de diferentes modelos de probabilidade. Aqui é digno de nota que a opção em alguns cursos de graduação pela redução ou mesmo supressão do estudo de probabilidade básica, leva a formação de profissionais que terão dificuldades para desenvolver análises estatísticas com enfoque nos modelos probabilísticos coerentes.

Como neste artigo tem-se como um dos objetivos a exemplificação de maneira simples do funcionamento da abordagem bayesiana para problemas ecológicos, analisaremos um caso em que se faz uso de famílias conjugadas. Mostraremos a partir do exemplo as etapas para a construção e exploração da distribuição *posterior*, incluindo pontos a serem observados que podem facilitar a identificação dos modelos de probabilidade adequados. Recomenda-se o livro de Gelman *et al.* (1995) para estudos e entendimento mais aprofundado dos elementos envolvidos na inferência bayesiana.

Para facilitar comparações, um mesmo problema ecológico será abordado de três maneiras diferentes:

a) Abordagem ortodoxa usualmente ensinada nos cursos de graduação;

b) Abordagem ortodoxa baseada na análise da verossimilhança;

c) Abordagem bayesiana em que se faz uso da verossimilhança e de uma distribuição *a priori*.

A ESTIMATIVA DE PROPORÇÕES: ESTUDO DA COBERTURA DE ALGAS MARINHAS EM UM COSTÃO ROCHOSO

Apresentação do problema e dos dados experimentais

Suponha que foi realizada uma pesquisa para o estudo da cobertura de algas marinhas em um costão rochoso. A amostragem foi feita com armações quadradas com área interna de 400 cm^2 (20 cm x 20 cm de lado) em que se verifica qual é a alga dominante, no sentido de que entre todas as espécies ela é a que recobre a maior área dentro do quadrado. Para cada quadrado, uma espécie de alga é a "dominante". Trinta destes quadrados amostrais foram distribuídos aleatoriamente em um costão rochoso verificando em cada um deles se há o domínio de uma particular espécie de alga verde. Suponha que estamos particularmente interessados em fazer inferências sobre a proporção de quadrados que cobrem a totalidade do costão rochoso em que há domínio desta espécie de alga verde. Vamos denominar de θ a proporção de quadrados de 400 cm^2 do costão em que predomina a alga de interesse.

No problema de inferência proposto temos duas metas:

a) estimar a proporção da superfície do costão dominada pela alga (θ); e

b) avaliar se a proporção dominada pela alga de interesse é maior que 50% ($\theta > 0{,}5$), ou seja, avaliar se a alga é dominante na maioria da superfície do costão.

Dois comentários são importantes aqui. O primeiro é que tais problemas surgem naturalmente em estudos ecológicos, e que o exemplo não é um caso pouco realista. A título de exemplificação podemos supor por exemplo que este costão rochoso é o único da área de estudo voltado para o norte, ou é um no qual não há muitas horas de insolação, e estaríamos tentando avaliar se nesta situação específica ocorre uma dominância clara da alga de interesse. O segundo comentário é sobre o procedimento utilizado no exemplo para se identificar "dominância" com os quadrados amostrais. Ela é só ilustrativa! Outras alternativas, talvez mais coerentes sob o prisma de um ecólogo, poderiam ser levantadas. O importante é ter em mente que qualquer que fosse a abordagem, ela poderia ser colocada no contexto de um problema de inferência estatística.

Voltando ao caso proposto para exemplificação poderíamos ter como resultado do programa amostral o número de quadrados amostrados ($n = 30$) e de quadrados amostrados em que há dominância da alga verde ($x = 18$).

Abordagem ortodoxa usualmente ensinada em cursos de graduação

Nos cursos de graduação e mesmo pós-graduação relacionados às ciências naturais, esta é a abordagem mais utilizada. A propagação destes ensinamentos resulta que a esmagadora maioria de artigos da área de ciências naturais realiza inferências utilizando as soluções descritas mais abaixo. Como esta abordagem é bem conhecida, não são feitos comentários mais detalhados. Livros textos de estatística básica (*e.g.* Bussab & Morettin, 2004; Triola, 2005) podem ser consultados pelos interessados. De maneira resumida pode-se dizer que na abordagem ortodoxa a inferência é realizada baseada na visão frequentista do significado de probabilidade e na concepção de que essa probabilidade é calculada considerando-se o dado observado como condicionado a um determinado valor assumido para o parâmetro de interesse como a proporção do nosso exemplo.

Na abordagem ortodoxa, para o cumprimento da primeira meta proposta no problema (obter estimativas de θ), poderíamos fazer uso do cálculo dos tradicionais intervalos de confiança para as proporções. O primeiro passo seria utilizar um estimador[1] ($\hat{\theta} = x/n$) para a proporção de quadrados com dominância da alga verde que neste caso seria $\hat{\theta} = 18/30 = 0{,}6$. Prosseguimos com a escolha do nível de confiança (usualmente 95%) da estimativa intervalar denominada "intervalo de confiança". Os quadrados amostrados são classificados em duas categorias ("dominado pela alga" e "não dominado pela alga"). Como a amostra de quadrados é tomada sem reposição, o número de quadrados amostrados com dominância da alga segue uma distribuição hipergeométrica, a qual pode justificadamente ser aproximada por uma binomial quando o número de quadrados amostrais (n) é muito menor que o número de quadrados da população (N), que no nosso caso corresponde ao costão inteiro. Nos casos em que $n \geq 30$ e em que se espera que o número de observações em ambas as categorias ("quadrados dominados" e "quadrados não dominados pela alga" no nosso caso) seja maior que 5, justifica-se por sua vez a aproximação da binomial pela distribuição normal, que é tradicionalmente a mais estudada em cursos de estatística. Assim nas abordagens ortodoxas o cálculo do intervalo de confiança para uma proporção θ, é realizado com o auxílio de uma variável (Z) que tem uma distribuição normal padronizada. Temos então que:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}} \tag{3}$$

Para a obtenção da solução substitui-se o parâmetro (θ) pelo estimador ($\hat{\theta}$) e tem-se então:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \tag{4}$$

Assim, o intervalo de confiança para a proporção θ é:

$$IC_{95\%}^{\theta} \Rightarrow \left]0{,}425;0{,}775\right[\tag{5}$$

Cabe lembrar que este intervalo de confiança significa que se o experimento pudesse ser e fosse repetido um grande número de vezes, poderia ser então construído um intervalo de confiança diferente em cada repetição. Noventa e cinco porcento destes intervalos devem conter o valor verdadeiro da proporção estudada caracterizando assim a "confiança" associada ao único intervalo de que dispomos na prática. Portanto, poderia ser assumido que qualquer um dos intervalos de confiança

[1] A justificativa para a escolha deste estimador aparece mais adiante quando se trata do estimador de maxima verossimilhança.

construídos, particularmente o que está mostrado acima, tem 95% de probabilidade de conter o valor da verdadeira proporção populacional (θ). No entanto, a partir de um intervalo de confiança não há indicações sobre os valores mais prováveis para o parâmetro. Por exemplo, ao avaliarmos o intervalo de confiança acima é incorreto pensar que o valor mais provável de θ é exatamente o ponto médio entre 0,425 e 0,775, que é o estimador $\hat{\theta} = 0{,}6$.

Vejamos agora como ficaria o teste de hipótese. Teríamos que estabelecer a hipótese de interesse como uma alternativa (H_1) à hipótese nula (H_0):

(6) $$\begin{cases} H_0 : \theta = 0{,}5 \\ H_1 : \theta > 0{,}5 \end{cases}$$

A estatística de teste é:

(7) $$z = \frac{\hat{\theta} - \theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \approx 1{,}095$$

Portanto, se é verdade que a proporção do costão recoberta pela alga é 0,5 (H_0), a probabilidade de que obtivéssemos uma amostra que resultasse em um valor de z tão ou mais extremo que 1,095 é $P(Z \geq 1{,}095) \approx 0{,}137$. Se adotarmos um nível de significância de $\alpha = 0{,}05$ vemos que a probabilidade indicada acima não é muito baixa pois $P(Z \geq 1{,}095) > \alpha$. A conclusão seria que não há nos dados evidências contrárias suficientes para que se rejeite H_0. Portanto, não há motivações para apoiar a hipótese de que a alga recobre a maioria da superfície do costão ($\theta > 0{,}5$).

A probabilidade de que a colocação feita acima esteja equivocada, ou seja, que a hipótese H_0 não tenha sido rejeitada mesmo sendo falsa, é a probabilidade de se estar cometendo um erro do tipo II, $P(erroII) \approx \beta$. O "poder de um teste" é a probabilidade de corretamente rejeitarmos H_0 quando ela é de fato falsa, que é $1-\beta$. Raramente é realizado um estudo do poder de teste nos artigos científicos em que se faz uso da abordagem ortodoxa, embora os estatísticos ortodoxos reconheçam que seja recomendável fazê-lo sempre que H_0 não é rejeitada. Na verdade, há a percepção equivocada de que quando a hipótese nula não é rejeitada o estudo é pouco importante e não tem apelo para a publicação. No entanto, mesmo que H_0 não seja rejeitada, o trabalho pode vir a ser relevante se houver um estudo do poder de teste.

O fato de que a hipótese nula de igualdade não foi rejeitada não quer dizer que ela é verdade. Por isso até é recomendável utilizar terminologias como "não rejeitamos H_0" ao invés de "aceitamos H_0". Enquanto a primeira colocação passa a impressão mais coerente de que H_0 é plausível mas não necessariamente verdade, a segunda passa a impressão de que H_0 deve ser tomada como uma verdade, o que não é o caso.

É importante salientar também que ao final do teste de hipótese disposto acima não se pode responder nem mesmo de maneira aproximada a perguntas de real interesse ecológico como por exemplo: "Qual é a probabilidade de que a proporção do costão recoberta pela alga seja maior que 0,5?". Com base no teste de hipótese acima simplesmente podemos dizer que os dados observados seriam razoavelmente plausíveis caso a hipótese nula ($\theta = 0{,}50$) fosse de fato verdade. No entanto, não estamos dizendo nada especificamente sobre a intensidade com que uma hipótese (nula ou alternativa) é mais plausível que a outra.

Abordagem baseada verossimilhança

Esta abordagem é similar à ortodoxa no que diz respeito à conceituação de probabilidade. No entanto, aqui, a exemplo do que ocorre também na abordagem bayesiana (mostrada mais abaixo), considera-se que a inferência para θ deve ser feita condicionada aos dados observados. Aos leitores interessados no estudo do princípio da verossimilhança, sugere-se a consulta de livros texto como o de Edwards (1972).

A título de exemplificação, vamos considerar que a distribuição binomial é razoável para a análise do problema de proporções. A função de verossimilhança para esta distribuição é:

(8) $p(dados \mid \theta) = L(x \mid \theta) = \binom{n}{x} \cdot \theta^x \cdot (1-\theta)^{n-x}$

Na verdade na função de verossimilhança o parâmetro dado está condicionado aos valores fixados de x e n. Para encontrar a melhor estimativa da proporção temos que obter o estimador de máxima verossimilhança $\hat{\theta}_{EMV}$. Para isto basta derivar a equação acima e igualar o resultado a zero. Fazendo isto vemos que $\hat{\theta}_{EMV} = x/n$.

A função acima (eq. 8) pode ser calculada para qualquer estimativa de θ no intervalo [0,1], incluindo $\hat{\theta}_{EMV}$, e pode-se obter uma curva com o perfil da verossimilhança. Para o cálculo do intervalo de confiança para θ podemos fazer uso de resultados teóricos (Hudson, 1971; Kendall & Stuart, 1979) que implicam que (-2) vezes o logaritmo da razão entre as verossimilhanças de um valor qualquer de θ e de $\hat{\theta}_{EMV}$,

(9) $$R = -2\log\frac{L(x \mid \theta)}{L(x \mid \hat{\theta}_{EMV})}$$

tem uma distribuição qui-quadrado (χ^2) com um grau de liberdade. Desta densidade de probabilidade temos que $P(\chi^2 < 3{,}481) = 0{,}95$. Então depois de calculado $\hat{\theta}_{EMV}$, basta encontrar os valores de θ que correspondem a $R = 3{,}841$, e teremos um intervalo de confiança (95%) aproximado, que no caso é

(10) $$IC_{95\%}^{\theta} \Rightarrow \left]0{,}422;0{,}762\right[$$

Uma ilustração deste procedimento é mostrada na figura 1.

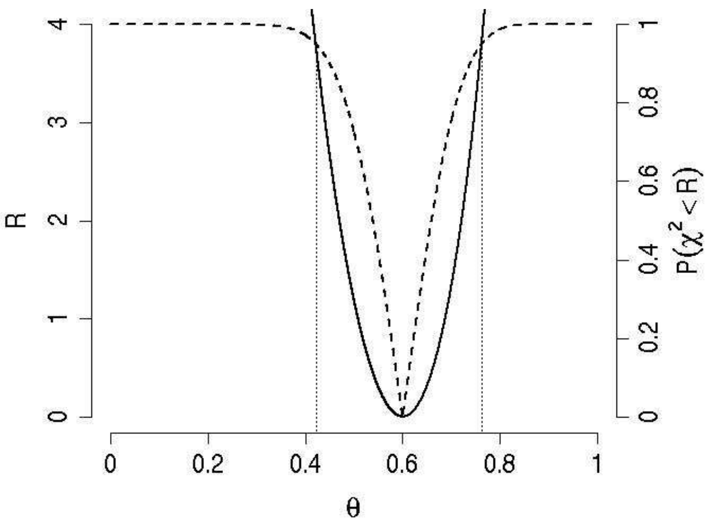

FIGURA 1 – Avaliação do perfil de verossimilhanças para cálculo de intervalos de conficança. A linha contínua (R) é menos duas vezes o logaritimo da razão entre verossimilhanças. A linha tracejada espessa é a probabilidade de que uma estatística de qui-quadrado (gl=1) seja menor do que R, $P(\chi^2 < R)$. A linha pontilhada extende-se desde o eixo que representa o parâmetro de interesse (θ) até o ponto de intersecção entre as duas outras curvas para $P(\chi^2 < R) = 0{,}95$. Os valores marcados no eixo θ correspondem ao intervalo de confiança (95%).

A solução mostrada na equação 9 também pode ser usada para se fazer testes entre duas hipóteses concorrentes, ou seja, valores alternativos para θ. Para dois valores quaisquer se a razão R não alcança 3,841, significa que não há uma evidência muito forte a favor do θ do denominador se comparado ao θ do numerador. Por exemplo se fizermos $\theta_{denominador} = 0{,}55$ e $\theta_{numerador} = 0{,}50$ temos que $R = 0{,}902$, o que implica que a hipótese 0,55 tem mais suporte que 0,50, mas, a diferença entre elas não é significativa (a um nível de 5% de significância). Fazendo uso deste princípio vemos que se comparado a um numerador de 0,5 não há um valor de θ que torne R maior que 3,841. Portanto, tecnicamente, não há uma hipótese para a proporção superior a $\theta = 0{,}5$ que seja significantemente melhor. Então, com a abordagem da verossimilhança podemos dizer que alguns valores de $\theta > 0{,}5$ têm mais suporte que $\theta = 0{,}5$ mas não se pode medir a diferença entre a credibilidade das diferentes hipóteses. Novamente não podemos responder a perguntas como: "Qual é a probabilidade de que a proporção do costão recoberta pela alga seja maior que 0,5?".

Abordagem Bayesiana

A função de verossimilhança é a mesma da equação 8. No entanto, agora a função de verossimilhança deve ser combinada com a distribuição *a priori* para se obter a distribuição *posterior* atualizada. Para a solução do problema proposto vamos utilizar duas distribuições *a priori*, uma não-informativa (ou difusa) e outra informativa.

a) Solução com uma distribuição *a priori* não-informativa

Para assumir uma ignorância completa sobre a proporção poderíamos assumir como *a priori* uma distribuição uniforme no intervalo [0,1] que é um caso particular de uma distribuição $Beta(a,b)$ (eq. 11) em que os parâmetros são $a = 1$ e $b = 1$ (ver linha pontilhada na figura 2 A).

(11) $p(\theta) = Beta(a,b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$

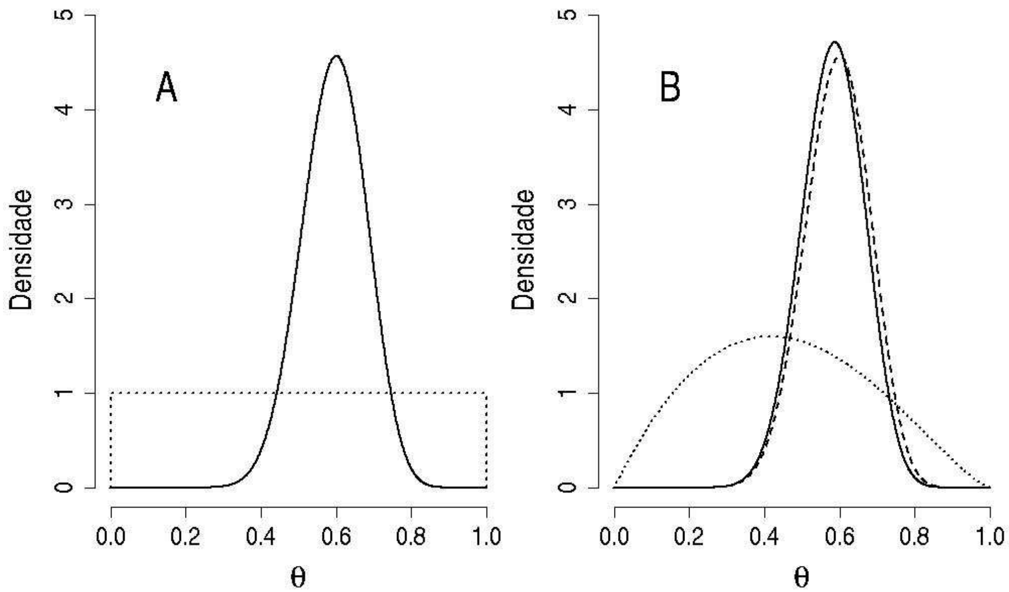

FIGURA 2 – Distribuições *a priori* (linha pontilhada), *posterior* (linha contínua) e a verossimilhança (linha tracejada) para as análises com *a priori* não-informativa (A) e informativa (B). No painel A há uma sobreposição entre a verossimilhança e a distribuição *posterior*. Todas as curvas estão normatizadas de tal forma que a área abaixo de cada uma delas integra a 1.

A escolha da distribuição Beta é conveniente uma vez que ela é conjugada da função de verossimilhança. Note como os dois últimos termos à direita da distribuição *a priori* (eq. 11) e da verossimilhança (eq. 8) são similares. O termo remanescente é uma constante e não envolve a quantidade de interesse θ.

De acordo com a equação 2, prosseguimos então com a multiplicação da distribuição *a priori* pela verossimilhança e obtemos:

(12) $p(\theta \mid dados) = p(\theta \mid x) \propto \theta^{a+x-1}(1-\theta)^{b+n-x+1}$

Com a constante já suprimida vemos que a equação acima corresponde a uma distribuição $Beta(a' = a + x, b' = b + n - x)$. Se lembrarmos que os parâmetros da distribuição *a priori* não-informativa são $a = b = 1$ e que amostramos $n = 30$ quadrados e que verificamos que $x = 18$ eram dominados pela alga de interesse, temos que a distribuição *posterior* para θ é $Beta(a' = 19, b' = 13)$ (figura 2 A).

Podemos usar então as propriedades da distribuição $Beta$ (ver Gelman *et al.*, 1995) para obtermos a esperança (média) $E(\theta)$, a variância $\mathrm{var}(\theta)$, e a moda $\mathrm{mod}(\theta)$ da distribuição *posterior* da proporção do costão recoberta pela alga:

(13) $$E(\theta) = \frac{a'}{a'+b'} \approx 0{,}598$$

$$\mathrm{var}(\theta) = \frac{a'b'}{(a'+b')^2(a'+b'+1)} \approx 0{,}007$$

$$\mathrm{mod}(\theta) = \frac{a'-1}{a'+b'-2} \approx 0{,}633$$

Um intervalo de 95% de confiança Bayesiano para θ pode ser obtido com o cálculo dos percentis de 2,5% e 97,5% da distribuição de densidade *posterior* $Beta(19,13)$:

(14) $$IC^{\theta}_{95\%} \Rightarrow \left]0{,}422;0{,}755\right[$$

Note que neste caso a solução é similar àquela obtida com a estatística ortodoxa, mas há uma diferença conceitual fundamental. Aqui podemos dizer que há 95% de probabilidade de que a proporção do costão recoberta pela alga seja um valor entre 0,422 e 0,775. Afirmações dessa natureza não podem ser feitas para os intervalos de confiança calculados com as outras abordagens. É importante mencionar que com a abordagem bayesiana pode-se inclusive indicar a moda de 0,633 como sendo o valor mais provável de θ, e mais do que isto, deve-se ressaltar que este valor mais provável geralmente não é o centro do intervalo de confiança.

Vamos agora tentar responder ao teste de hipótese proposto originalmente: "A proporção recoberta pela alga é maior que 0,5 (50%)?" Essa questão pode ser abordada de maneira direta, é só calcular a área da distribuição de probabilidade *posterior* $Beta(19,13)$, que corresponde a valores de $\theta > 0{,}5$. Obtemos então $p(\theta > 0{,}5 \mid dados) \approx 0{,}86$ e o complementar $p(\theta \leq 0{,}5 \mid dados) \approx 0{,}14$. Diríamos então que há 86% de probabilidade de que θ seja maior que 0,5, e que as evidências são em muito favoráveis (*i. e.* as chances a favor de $\theta > 0{,}5$ são aproximadamente de 6 para 1) à hipótese de que a proporção do costão recoberta seja realmente maior que 50%.

É oportuno neste ponto chamar a atenção do leitor para o fato de que nenhuma das hipóteses contrapostas na abordagem bayesiana incluem exclusivamente uma igualdade, o que no caso de θ não faz realmente sentido. A probabilidade para qualquer igualdade no estudo de uma proporção (θ), que é uma variável contínua, seria sempre igual a zero. Por exemplo, no nosso caso teríamos que $p(\theta = 0{,}5 \mid dados) = 0$. Este elemento expõe um pouco mais a dificuldade de se lidar com os testes de hipóteses ortodoxos onde se começa com algo como "Suponhamos que $\theta = 0{,}5$ é verdade, então ...", quando na realidade a probabilidade de que isto seja verdade é zero! Vamos voltar agora à abordagem bayesiana, com a avaliação da solução para a outra distribuição *a priori*.

b) Solução com uma distribuição *a priori* informativa

Vamos considerar agora que há alguns estudos prévios que indicam que a proporção da superfície recoberta pela alga em costões similares àquele que estamos estudando é de 0,42 (ou 42%). Não seria lógico desconsiderar deliberadamente esta informação. Uma estimativa coerente para a proporção deveria ser uma combinação do que já se sabia (*a priori*) com o que aprendemos com o experimento atual (verossimilhança). O primeiro passo é a construção da distribuição *a priori*.

Há diversas maneiras e mesmo regras sugeridas por alguns autores sobre os procedimentos que se deve tomar para aferir os parâmetros de distribuições *a priori* a partir de informações pretéritas como as dispostas no início do parágrafo acima. Berry (1995) discute especificamente como construir distribuições *a priori* para proporções com o uso de distribuições $Beta$. Aqui utilizaremos um procedimento bem simples. A moda para um valor θ que se distribui de acordo com uma distribuição $Beta(a,b)$ é:

(15) $$\mathrm{mod}(\theta) = \frac{a-1}{a+b-2}$$

A questão então é encontrar pares de valores para a e b que resultem em uma moda igual a 0,42 (informação disponível). Quanto menores forem os valores de a e b, mais aberta e menos informativa será a curva de densidade da distribuição. Optamos então por usar uma distribuição $Beta(a = 2, b = 2{,}4)$ como exemplo (linha pontilhada na figura 2 B).

O procedimento para a obtenção da distribuição *posterior* é idêntico ao usado mais acima quando consideramos o uso de uma distribuição *a priori* não-informativa. Multiplicando a verossimilhança pela distribuição *a priori*, obtemos novamente uma solução (eq. 12) em que reconhecemos que a *posterior* é agora a distribuição $Beta(a + x = 20, b + n - x = 14{,}4)$. A esperança (média) $E(\theta)$, a variância $\text{var}(\theta)$, e a moda $\text{mod}(\theta)$, para a proporção do costão recoberta pela alga são:

(16) $E(\theta) \approx 0{,}581 \; \text{var}(\theta) \approx 0{,}07 \; \text{mod}(\theta) \approx 0{,}586$

O intervalo de 95% de confiança Bayesiano para θ é

(17) $$IC_{95\%}^{\theta} \Rightarrow \left]0{,}416;0{,}738\right[$$

e portanto, há 95% de probabilidade de que a proporção do costão recoberta pela alga seja um valor entre 0,416 e 0,738. O valor de θ mais provável é 0,586, o qual novamente não corresponde ao centro do intervalo de confiança.

Com a exploração da distribuição *posterior* obtemos que $P(\theta > 0{,}5) \approx 0{,}83$ e diríamos que há 83% de probabilidade de que θ seja maior que 0,5. Portanto as evidências a favor de $\theta > 0{,}5$ diminuem (*i.e. as* chances a favor de $\theta > 0{,}5$ se reduzem para aproximadamente 5 para 1), mas continuam sendo em muito favoráveis à hipótese original de que a proporção do costão recoberta pela alga é realmente maior que 50%.

Um sumário dos resultados obtidos na análise dos dados quando utilizados os três diferentes procedimentos para a inferência consta na tabela 1. Exposto isto podemos avançar para uma comparação dos resultados e para as considerações finais.

TABELA 1 – Intervalos de confiança e resultados dos testes de hipóteses para a proporção do costão recoberta pela alga marinha.

Procedimento		Intervalo de 95% de confiança	Avaliação da hipótese de que a alga recobre a maioria do costão ($\theta > 0{,}5$)
Abordagem ortodoxa		$\left]0{,}425;0{,}775\right[$	Não há uma resposta direta. Não podemos rejeitar que $\theta = 0{,}5$ e portanto não há como dar suporte para a hipótese de que $\theta > 0{,}5$ ainda que aparentemente ela seja melhor que $\theta = 0{,}5$.
Verossimilhança		$\left]0{,}422;0{,}762\right[$	Não há uma resposta direta. A hipótese de que $\theta = 0{,}5$ é pior que algumas outras em que $\theta > 0{,}5$ mas não o suficiente para abdicarmos de $\theta = 0{,}5$ em favor de qualquer outra hipótese.
Bayesiano	*a priori* não-informativa	$\left]0{,}422;0{,}755\right[$	Há uma probabilidade de 0,86 (86%) de que a alga cubra a maioria do costão, ou seja, de que $\theta > 0{,}5$, e portanto esta é a hipótese mais plausível.
	a priori informativa	$\left]0{,}416;0{,}738\right[$	Há uma probabilidade de 0,83 (83%) de que a alga cubra a maioria do costão, ou seja, de que $\theta > 0{,}5$, e portanto esta é a hipótese mais plausível.

DISCUSSÃO

Parte da diferença encontrada ao se comparar os intervalos de confiança da abordagem ortodoxa com as demais (verossimilhança ou bayesiana) se deve ao fato de que naquela se faz uma aproximação pela Normal. De certa forma o uso de aproximações com uso de tabelas já disponíveis é justificado quando amostras são grandes e/ou a distribuição *posterior* envolvida é unimodal e razoavelmente simétrica. Provavelmente também é a melhor opção para cursos onde aulas em salas com computadores não possam ser realizadas com facilidade. Mas acreditamos que isto está mudando rapidamente. Hoje em dia, pacotes computacionais básicos incluem funções estatísticas que permitem o estudo e cálculos pertinentes para grande parte das distribuições de probabilidade de uso comum, como é o caso da binomial.

A simplicidade da abordagem e os cálculos que não exigem ferramentas computacionais são certamente as maiores vantagens do procedimento que envolve a aproximação da binomial por uma normal. No entanto, na sua aplicação há uma violação conceitual, pelo menos no que se refere à replicabilidade do evento estudado, que nem sempre é aplicável em problemas ecológicos. Independentemente de haver ou não violações internas dos preceitos, é importante notar que o teste de hipótese realizado com a abordagem ortodoxa, ensinado nos cursos de graduação, é insatisfatório pois não permite a elaboração de uma resposta direta às questões usualmente propostas.

O ponto crucial referente à abordagem com o emprego da verossimilhança é que dado os resultados do intervalo de confiança e do teste de hipótese não se atribuem probabilidades sobre quais seriam os plausíveis valores do parâmetro de interesse θ, que no exemplo apresentado é a proporção do costão rochoso dominada pela alga marinha. Podemos por exemplo dizer que a hipótese $\theta = 0{,}55$ tem mais suporte que a hipótese $\theta = 0{,}5$, mas não podemos mensurar o quão $\theta > 0{,}5$ é mais plausível. Essa impossibilidade resulta em uma frustração do pesquisador ao final do trabalho.

A abordagem via verossimilhança é de certa forma muito próxima da bayesiana. Na verdade as curvas de verossimilhança não normatizadas (*i.e.* não escalonadas para integrar a 1) contém informações dos dados sobre os pesos das evidências em favor de uma ou outra hipótese. A diferença é que há algumas restrições tomadas como imprescindíveis pelos adeptos da verossimilhança e que são desnecessárias no contexto bayesiano. A primeira está na recusa em aceitar probabilidades fora do paradigma freqüentista. A segunda, conseqüência direta da primeira, está na recusa de combinar verossimilhança com a distribuição *a priori*. Se um pesquisador adota uma noção mais geral de probabilidade como quantificador lógico, não há maiores dificuldades em aceitar verossimilhanças como um caso particular (e incompleto) de uma distribuição *posterior*.

Nota-se que não há uma distinção grande quanto aos intervalos de confiança, principalmente entre aqueles obtidos com a verossimilhança, e com a abordagem bayesiana em que se fez uso de uma distribuição *a priori* não informativa. Na verdade quando se adota uma *a priori* desta natureza a distribuição *posterior* acaba sendo delineada fundamentalmente pelo dado. Em outras palavras, o dado “fala por si mesmo” a partir da função de verossimilhança, e a distribuição a *priori* tem pouco efeito no resultado final. Portanto, na maioria dos casos a análise baseada somente na verossimilhança, e a abordagem bayesiana com uma distribuição *a priori* não-informativa, acabam resultando em limites similares de intervalos de confiança. Como mencionado anteriormente a grande diferença entre estas duas abordagens é conceitual e interfere na interpretação do que significam os intervalos de confiança calculados.

Na abordagem bayesiana, podem ser naturalmente encontradas respostas a questões de timbre ecológico, sem que se incorra em uma violação interna dos preceitos da abordagem. Normalmente as críticas que provêm de estatísticos ortodoxos são exatamente sobre alguns dos preceitos intrínsecos da abordagem bayesiana, como por exemplo, o tratamento da probabilidade como o grau de plausibilidade de uma proposição, e o uso de distribuições *a priori* de teor subjetivo. A avaliação da primeira destas críticas envolve desafios conceituais bastante técnicos no que diz respeito a operacionalização de lógica indutiva. Para os interessados, uma abordagem introdutória pode ser encontrada em Kinas & Andrade (2007). Tratamento aprofundado do tema encontra-se em Cox (1946), Pólya (1954) e no excelente livro de Jaynes (2003).

No que tange a crítica de que a distribuição *a priori*, como elemento subjetivo, macula uma abordagem científica, simplesmente alertamos que a própria verossimilhança, cerne da abordagem ortodoxa, é igualmente subjetiva. Finalmente, no que toca a contenda entre as abordagens ortodoxas e bayesianas, entendemos que esta última é, no momento, a única que produz respostas diretas a perguntas relevantes no contexto ecológico e, portanto, há um grande apelo para o seu uso.

REFERÊNCIAS

BAYES, Rev. T. 1763. An essay toward solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Lond.*, 370-418.

BEGER, JO. 1985. Statistical Decision Theory and Bayesian Analysis. New York, Springer-Verlag. 617p.

COX, RT. 1946. Probability, frequency, and reasonable expectation. *Am. J. Physic.*, 14: 1-13.

EDWARDS, AWF. 1972. Likelihood. London, The John Hopkins University Press. 275 p.

GELMAN, A, JB CARLIN, HS STERN & DB RUBIN. 1995. Bayesian Data Analysis. London, Chapman & Hall. 526p.

HUDSON, DJ. 1971. Interval estimation from the likelihood function. *J. R. Statistic. Soc.*, 33: 256-262.

JAYNES, ET. 2003. Probability Theory: The Logic of Science. New York, Cambridge University Press. 727p.

KENDALL, M & A STUART. 1979. The advanced theory of statistics. New York, Oxford University Press. 758p.

KINAS, PG & HA ANDRADE. 2007. Bayesian statistics for fishery stock assessment and management: a synthesis. *PANAMJAS*, 2(2): 103-112.

PÓLYA G. 1954. Mathematics and Plausible Reasoning. (2 vls). Princeton, Princeton University Press. 296p.

R Development Core Team. 2007. R: A language and environment for statistical computing. www.R-project.org.

SALSBURG, D. 2002. The Lady Tasting Tea: How Statistics Revolutionized Science in Twentieth Century. New York, Owl Books. 352p.

Entrada: 12/11/2007
Aceite: 30/06/2008