UNIVERSIDADE FEDERAL DE ALFENAS
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS - ICSA

LUCAS BARATA ALVES MACHADO

INFERÊNCIA BAYESIANA: UMA REVISÃO E ESTUDO DE SIMULAÇÃO

VARGINHA/ MG
2024

LUCAS BARATA ALVES MACHADO

INFERÊNCIA BAYESIANA: UMA REVISÃO E ESTUDO DE SIMULAÇÃO

Trabalho apresentado como parte dos requisitos para obtenção do título do Bacharelado interdisciplinar de Ciência e Economia pela Universidade Federal de Alfenas.

Orientadora: Profª. Drª. Gislene Araújo Pereira

VARGINHA / MG

2024

LUCAS BARATA ALVES MACHADO

INFERÊNCIA BAYESIANA: UMA REVISÃO E ESTUDO DE SIMULAÇÃO

Trabalho apresentado como parte dos requisitos para obtenção do título do Bacharelado interdisciplinar de Ciência e Economia pela Universidade Federal de Alfenas.

Orientadora: Profª. Drª. Gislene Araújo Pereira

Aprovado em:

Profª. Drª. Gislene Araújo Pereira

Universidade Federal de Alfenas

Prof°. Dr°. Manoel Vitor de Souza Veloso

Universidade Federal de Alfenas

Profª. Drª. Leticia Lima Milani Rodrigues

Universidade Federal de Alfenas

AGRADECIMENTO

Gostaria de expressar minha profunda gratidão à minha orientadora, Gislene. Sem sua orientação e dedicação, este trabalho não teria sido possível. Agradeço por toda a paciência, suporte e incentivo ao longo desta jornada.

Agradeço também ao meu amigo Luis Heitor, meu querido colega de estudos e companheiro de desafios. Sua amizade e colaboração foram essenciais para a realização deste trabalho.

Por fim, agradeço aos meus amigos Guilherme e Igor, por sempre estarem ao meu lado, não importa qual o problema. Vocês foram um suporte inestimável durante todo este processo.

RESUMO

Este trabalho apresenta uma revisão de literatura sobre a inferência bayesiana, uma abordagem poderosa na análise estatística moderna. A inferência bayesiana, diferentemente da abordagem frequentista tradicional, incorpora probabilidades subjetivas e conhecimento prévio para atualizar crenças sobre parâmetros desconhecidos com base em novas evidências. Esta metodologia é particularmente útil em cenários onde a disponibilidade de dados é limitada ou quando é crucial considerar informações prévias. Além da extensa revisão foram realizados dois estudos de simulação que exemplificam o uso das prioris conjugadas para estimar a probabilidade de sucesso em experimentos binomiais e a aplicação do amostrador de Gibbs em um modelo de regressão linear simples. Por fim, as considerações finais abordam tanto as vantagens quanto as limitações da inferência bayesiana, destacando seu potencial em áreas como machine learning e modelagem preditiva, enquanto discute propostas para futuras pesquisas, incluindo o uso de técnicas avançadas de Métodos de Monte Carlo via Cadeias de Markov (MCMC) como o algoritmo de Metropolis-Hastings e o Monte Carlo Hamiltoniano. Este estudo não apenas reforça a importância crescente da inferência bayesiana na análise estatística contemporânea, como também destaca seu papel crucial na formulação de modelos robustos e adaptativos para problemas complexos em diversas disciplinas científicas e práticas aplicadas.

Palavras-chave: Inferência Bayesiana, Simulação, Prioris Conjugadas, Amostrador de Gibbs

ABSTRACT

This paper provides an review of Bayesian inference bibliography, a powerful approach in modern statistical analysis. Unlike traditional frequentist methods, Bayesian inference incorporates subjective probabilities and prior knowledge to update beliefs about unknown parameters based on new evidence. This methodology is particularly useful in scenarios with limited data availability or when considering prior information is crucial. In addition to the comprehensive review, two simulation studies were conducted to illustrate the use of conjugate priors for estimating success probabilities in binomial experiments and the application of the Gibbs sampler in a simple linear regression model. Finally, the concluding remarks discuss both the advantages and limitations of Bayesian inference, highlighting its potential in areas such as machine learning and predictive modeling, while proposing avenues for future research, including advanced Markov Chain Monte Carlo MCMC techniques like the Metropolis-Hastings algorithm and Hamiltonian Monte Carlo. This study not only underscores the growing importance of Bayesian inference in contemporary statistical analysis but also emphasizes its critical role in developing robust and adaptive models for complex problems across various scientific disciplines and applied practices.

Keywords: Bayesian Inference, Simulation, Conjugate Priors, Gibbs Sampler

SUMÁRIO

1 INTRODUÇÃO

A análise estatística desempenha um papel fundamental em diversas áreas do conhecimento, permitindo a coleta, organização, interpretação e apresentação de dados. Ela se divide em duas grandes categorias: a análise descritiva e a análise inferencial. Enquanto a análise descritiva se concentra em resumir e descrever os dados coletados, a análise inferencial vai além, utilizando esses dados para fazer previsões ou inferências sobre uma população mais ampla.

A análise inferencial é um processo crítico na estatística, pois visa inferir parâmetros desconhecidos de uma população com base em dados amostrais. Em outras palavras, busca-se tirar conclusões sobre uma população com base em uma amostra representativa. A análise inferencial pode ser abordada de duas principais perspectivas: a frequentista e a bayesiana.

A inferência frequentista é a abordagem tradicional e mais amplamente utilizada na estatística. Ela se baseia na frequência ou proporção de eventos observados em amostras repetidas. A inferência frequentista faz uso de estimadores pontuais, intervalos de confiança e testes de hipóteses para tirar conclusões sobre a população (LUNN et al., 2012).

Por outro lado, a inferência bayesiana adota uma perspectiva diferente, incorporando a probabilidade subjetiva. A inferência bayesiana utiliza o Teorema de Bayes para atualizar a probabilidade de uma hipótese à medida que novas evidências são obtidas. Essa abordagem permite a inclusão de informações a priori, ou seja, conhecimentos prévios sobre o fenômeno estudado, para refinar as estimativas e inferências (BERGER,1985).

O histórico da inferência bayesiana remonta ao século XVIII, com os trabalhos do reverendo Thomas Bayes. No entanto, foi somente no século XX, com o avanço dos computadores e dos métodos numéricos, que a abordagem bayesiana ganhou maior destaque e aplicabilidade prática. A flexibilidade e a capacidade de incorporar conhecimentos prévios tornaram a inferência bayesiana uma ferramenta poderosa em diversas áreas, como biologia, medicina, economia e engenharia (PRESS,2003) .

Entre as principais vantagens da inferência bayesiana destaca-se a ausência de necessidade de pressuposições rígidas sobre a distribuição dos dados, algo que muitas vezes é exigido na inferência frequentista. Além disso, a inferência bayesiana permite uma interpretação mais intuitiva das probabilidades e a possibilidade de mensuração da incerteza

O objetivo deste trabalho é realizar uma revisão bibliográfica sobre a inferência bayesiana e conduzir estudos de simulação para demonstrar suas aplicações e vantagens. Através desta revisão e dos estudos de simulação, o presente estudo visa aprofundar a compreensão sobre as metodologias bayesianas e suas contribuições para a análise estatística moderna.

Este trabalho está estruturado em cinco seções principais. Inicia-se com a presente Introdução, seguido pelo Desenvolvimento, onde é realizada uma revisão detalhada da literatura sobre a inferência bayesiana, destacando seus princípios, vantagens e aplicações. Em seguida, uma descrição da metodologia utilizada. Posteriormente, são conduzidos estudos de simulação para que possamos observar na prática o funcionamento do amostrador de Gibbs e das prioris conjugadas. Por fim, nas Considerações Finais, são discutidas as conclusões e dificuldades do trabalho e sugestões para pesquisas futuras.

2 DESENVOLVIMENTO

A inferência Bayesiana é uma abordagem estatística que permite atualizar a probabilidade de uma hipótese à medida que novas evidências ou informações se tornam disponíveis. Baseia-se no teorema de Bayes, que formaliza o mecanismo de ajuste de nossas crenças à luz de novos dados, que de acordo com Gelman et.al.(2013) é formalizado da seguinte forma:

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)}$$

Em que temos:

•$P(H|E)$: é a probabilidade posterior da hipótese H dada a evidência E;
•$P(E|H)$: é a probabilidade da evidência E sob a hipótese H;
•$P(H)$: é a probabilidade anterior da hipótese H (antes de observar a evidência E);
•$P(E)$: é a probabilidade da evidência E.

2.1 INFERÊNCIA BAYESIANA VS INFERÊNCIA FREQUENTISTA

Para escolher qual abordagem utilizar, é essencial compreender as diferenças entre a inferência bayesiana e a inferência frequentista. Essa diferenciação pode ser feita com base nos princípios fundamentais de cada abordagem, conforme exposto por Casella e Berger (2013), e apresentados a seguir.

I- Princípios frequentistas:

- Probabilidade refere-se a frequências relativas no longo prazo. As probabilidades são propriedades objetivas do mundo real.
- Parâmetros são constantes fixas e desconhecidas. Como eles não flutuam, não se podem fazer declarações úteis de probabilidade sobre os parâmetros.
- Procedimentos estatísticos devem ser desenhados para ter propriedades de

frequência bem definidas no longo prazo. Por exemplo, um intervalo de confiança de 95% deve conter o valor verdadeiro do parâmetro com uma frequência de pelo menos 95% em repetições infinitas do experimento.

II- Princípios Bayesianos

- A probabilidade descreve o grau de crença, não a frequência limitante. Assim, pode-se fazer declarações de probabilidade sobre várias coisas, não apenas sobre dados sujeitos a variação aleatória. Por exemplo, pode se dizer que "a probabilidade de Albert Einstein ter bebido uma xícara de chá em 1º de agosto de 1948" é 0,35. Isso não se refere a nenhuma frequência limitante. Reflete minha força de crença de que a proposição é verdadeira.
- É possível fazer declarações de probabilidade sobre parâmetros, mesmo que sejam constantes fixas.
- Realizam-se inferências sobre um parâmetro θ produzindo uma distribuição de probabilidade para θ. Inferências, como estimativas pontuais e intervalos de credibilidade, podem ser extraídas dessa distribuição.

2.2 O MÉTODO BAYESIANO

No contexto bayesiano, o que se busca é uma maneira de modelar crenças prévias e as implicações dessas interações entre os fatores antes conhecidos e aqueles observados (KRUSCHKE, 2014).

Do ponto de vista de Gelman et.al. (2014), a inferência bayesiana oferece uma maneira elegante agregar conhecimento prévio com dados novos, permitindo a realização de inferências sobre parâmetros desconhecidos de forma contínua e dinâmica. Essa abordagem é particularmente poderosa em contextos nos quais o conhecimento prévio é valioso e onde a incerteza é intrínseca ao problema. A flexibilidade e a capacidade de atualizar crenças à medida que novas informações são adquiridas fazem da inferência bayesiana uma ferramenta essencial em muitas áreas da estatística moderna e da ciência de dados.

Dessa maneira é introduzida a função de verossimilhança. Essa seria a função da qual tem-se que seu produto pela priori é proporcional à posteriori, que seria o resultado da atualização da crença (priori) pela observação realizada (verossimilhança).

A constante de proporção é um componente essencial na inferência bayesiana, assegurando que a distribuição a posteriori seja uma distribuição de probabilidade válida. Ela integra a informação da priori e da verossimilhança, permitindo que atualizemos nossas crenças sobre os parâmetros de maneira consistente e robusta. A normalização proporcionada por essa constante é fundamental para a interpretação e comparação de modelos bayesianos.

Como na ótica escolhida nosso parâmetro passa a ser uma variável aleatória associada a ela há uma distribuição de probabilidade. Faz sentido agora modelarmos essa distribuição, baseado em nossa crença e conhecimento a respeito do fenômeno em questão, que descreva o comportamento aleatório do nosso parâmetro a ser estimado. Essa distribuição é conhecida como priori.

2.3 PRIORI

A introdução da distribuição a priori é a novidade trazida pela análise bayesiana em comparação com a análise clássica. Além disso, ela é o aspecto mais relevante e frequentemente alvo de críticas na análise bayesiana. Uma preocupação com essa distribuição é que ela deve refletir o conhecimento prévio sobre a quantidade desconhecida θ, antes de se fazerem observações. A determinação dessa distribuição é geralmente subjetiva, mas é possível utilizar dados observados anteriormente para defini-la (CASELLA e BERGER, 2013).

Nas Subseções a seguir, será abordado alguns dos principais tipos de priori.

2.3.1 Priori de Jeffreys

A priori de Jeffreys é uma escolha específica de distribuição a priori utilizada na inferência bayesiana, proposta por Harold Jeffrey em 1961. É projetada para ser não informativa, ou seja, para introduzir o mínimo de informação possível sobre o parâmetro desconhecido, a fim de que as conclusões sejam dominadas pelos dados observados. A priori de Jeffrey’s é derivada a partir da matriz de informação de Fisher e é invariável sob transformações reparametrizantes, o que a torna uma ferramenta útil e objetiva em diversas aplicações estatísticas. (JEFFREYS, 1961)

Segundo Batista (2021), temos:

$$\mathrm{P}(\theta) = \sqrt{I(\theta)}$$

Onde:

- Caso uniparamétrico $I(\theta)$, Informação de fisher é dado por:

$$I(\theta) = E\left[\left(\frac{\delta}{\delta\theta} ln \ f(x|\theta)\right)^2 |\theta\right]$$

- Caso multiparamétrico, a matriz de informação de Fisher é dada por **I(θ)**, na qual o ij-ésimo elemento da matriz é:

$$\boldsymbol{I_{ij}}(\boldsymbol{\theta}) = E\left(\frac{\delta lnf(x|\theta)}{\delta\theta_i} . \frac{\delta lnf(x|\theta)}{\delta\theta_j} |\theta\right)$$

2.3.2 Prioris não informativas

Uma "priori não informativa" é um tipo de priori que representa uma ausência de conhecimento pré-existente ou uma falta de informações específicas sobre o parâmetro em questão (GELMAN et al., 2013).

Basicamente, ao utilizar uma priori não informativa, assume-se que todos os valores possíveis do parâmetro têm a mesma probabilidade antes de analisarmos os dados. Isso permite que os dados coletados desempenhem o papel principal na definição das conclusões.

Por exemplo, se quisermos estimar a probabilidade de um evento, como a chance de um time ganhar um jogo, e não temos informações prévias sobre os times, podemos adotar uma priori não informativa. Isso significa assumir que a probabilidade de vitória pode ser qualquer valor entre 0% e 100%, todos com a mesma chance. À medida que obtemos dados (como resultados de jogos anteriores), esses dados começam a "informar" nossa estimativa da probabilidade.

A utilização de uma priori não informativa assegura que as análises sejam guiadas principalmente pelos dados observados, reduzindo a influência de suposições iniciais.

2.3.3 Prioris conjugadas

Uma "priori conjugada" é uma escolha especial da distribuição a priori que, quando combinada com a distribuição dos dados (a função de verossimilhança), resulta em uma distribuição posterior que pertence à mesma família da distribuição a priori. Em outras palavras, o núcleo da distribuição a priori e a distribuição posterior são os mesmos, o que simplifica muito os cálculos (BERNARDO e SMITH, 2000).

Por exemplo, imagine que o interesse seja estimar a proporção de votos que um candidato receberá em uma eleição. Se atribuir uma distribuição beta como priori para a proporção de votos e os dados observados seguem uma distribuição binomial (número de votos recebidos), a distribuição posterior resultante também será uma distribuição beta. Essa relação simplifica os cálculos, pois não se necessita usar métodos numéricos complexos para encontrar a posterior.

Usar prioris conjugadas é muito útil porque facilita a atualização das crenças com novos dados de maneira analítica e direta, mantendo a mesma estrutura matemática. Isso é especialmente valioso em situações onde atualizações rápidas e frequentes são necessárias.

> A partir do conhecimento que se tem sobre θ, pode-se definir uma família paramétrica de densidades. Neste caso, a distribuição a priori é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros indexadores da família de distribuições a priori são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse θ. (EHLERS, 2003, p. 10).

Ainda de acordo com Ehlers (2003), a ideia é que as distribuições a priori e a posteriori pertençam à mesma classe de distribuições, de modo que a atualização do conhecimento sobre θ envolva apenas uma mudança nos hiperparâmetros. Nesse cenário, o aspecto sequencial do método Bayesiano pode ser explorado definindo apenas a regra de atualização dos hiperparâmetros, já que as distribuições permanecem inalteradas.

Como citado acima, quando as distribuições a priori e a posteriori pertencem a uma mesma classe de distribuições, tem-se uma conjugação entre elas.

A seguir será apresentado um exemplo para o caso das distribuições Binomial e Beta.

Exemplo: Considere uma amostra X de tamanho n da distribuição de Bernoulli com parâmetro θ. A densidade conjunta da amostra será:

$$p(\boldsymbol{x}|\theta) = \theta^t(1- \theta)^{n-t} = l(\theta; \boldsymbol{x}), \text{ onde } t = \sum_{i=1}^{n} x_i \text{ e } x_i = 0{,}1,\ i = 1, \ldots, n$$

que define uma classe de distribuições parametrizadas por $\theta \in [0,1]$. Do teorema de Bayes tem-se que a densidade a posteriori de θ dado $\boldsymbol{x}$ será da forma:

$$p(\theta|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta)\, p(\theta) \Rightarrow$$

$$\Rightarrow p(\theta|\boldsymbol{x}) \propto \theta^t(1-\theta)^{n-t}p(\theta).$$

Observa-se que as distribuições a priori e a posteriori estão conectadas pela verossimilhança; sendo assim, uma forma de construir a distribuição conjugada é se basear no núcleo da verossimilhança que é da forma $\theta^a(1-\theta)^b$, que caracteriza a família das distribuições Beta. Com isso toma-se para θ uma priori Beta. Isto é,

$$p(\theta) = \frac{1}{B(a,\beta)}\theta^{a-1}(1-\theta)^{\beta-1}, 0 < \theta < 1 \text{ e } a, \beta > 0.$$

Combinando, a priori Beta com a verossimilhança tem-se

$$p(\theta|\boldsymbol{x}) = \frac{(\theta^t(1-\theta)^{n-t})(\frac{1}{B(a,\beta)}\theta^{a-1}(1-\theta)^{\beta-1})}{\frac{1}{B(a,\beta)}\int_0^1 \theta^t(1-\theta)^{n-t})\theta^{a-1}(1-\theta)^{\beta-1}d\theta} =$$

$$= \frac{\theta^{a+t-1}(1-\theta)^{n+\beta-t-1}}{\int_0^1(\theta^{a+t-1}(1-\theta)^{\beta+n-t-1}\, d\theta} =$$

$$= \frac{\theta^{a+t-1}(1-\theta)^{n+\beta-t-1}}{B(a+t, \beta+n-t)} \Rightarrow$$

$$\Rightarrow p(\theta|\boldsymbol{x}) \propto \theta^{a+t-1}(1-\theta)^{n+\beta-t-1}$$

Portanto, $\theta|\boldsymbol{x} \sim B(a+t, \beta+n-t)$ desta forma a família de distribuições Beta é conjugada à Bernoulli (e consequentemente à binomial). A constante de proporcionalidade, na densidade $\theta|\boldsymbol{x}$ é $\propto = \frac{1}{B(a+t,\beta+n-t)}$, conforme demonstrado acima.

2.4 FAMILIA EXPONENCIAL

De acordo com Ehlers (2003), em alguns casos o método de conjugar prioris com suas posteriori respectivas não está ligado à incerteza a respeito crença anterior, e sim à propriedades de manejo tratativo análitico ganhas com tal suposição.

Ainda segundo esse autor, a família exponencial abrange muitas das distribuições de probabilidade mais comuns em Estatística, incluindo tanto as contínuas quanto as discretas. Uma característica fundamental dessa família é que ela possui uma estatística suficiente com dimensão fixa. (EHLERS, 2003).

Definição: A família de distribuição com função de (densidade) de probabilidade f(x|θ) pertence à família exponencial, caso seja possível manipular para o seguinte formato:

$$f(x|\theta) = a(x)\exp\{u(x)\Phi(\theta) + b(\theta)\}I_A(x)$$

Segundo o critério de Neyman-Pearson u(x) é uma estatística suficiente para θ.

Exemplo: A função de densidade da distribuição normal é dada por:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Sabendo que o formato da família exponencial é:

$$f(x|\boldsymbol{\theta}) = a(x)\exp\{u(x)\Phi(\boldsymbol{\theta}) + b(\boldsymbol{\theta})\}\, I_A(x)$$

Para mostrar que a distribuição normal é membro da família exponencial precisamos reescrevê-la da seguinte maneira:

$$f(x|\mu,\sigma^2) = \exp\left\{\frac{-x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\frac{\mu^2}{2\sigma^2} - \frac{\log(2\,\pi\sigma^2)}{2}\right\}$$

Onde:

- $a(x) \ = 1$
- $\Phi(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2}$
- $u(x) = (x, x^2)$
- $b(\boldsymbol{\theta}) = \frac{\mu^2}{2\sigma^2} + \frac{\log(2\,\pi\sigma^2)}{2}$

2.5 FUNÇAO DE VEROSSIMILHANÇA

Para os dados observados $\boldsymbol{x}$, a funcão $L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}|\theta)$, considerada como função de θ, é chamada *função de verossimilhança* (BERGER,1985).

De acordo com (Berger 1985, pg.27), a intuição por trás do nome ‘função de verossimilhança’ (‘*likelihood function*’) é que mais verossímil (*likely*) é o θ quanto maior for $f(\boldsymbol{x}|\theta)$.

O princípio da verossimilhança afirma que o “significado evidencial” dos resultados experimentais é caracterizado integralmente pela função de verossimilhança, sem outra referência à estrutura de um experimento, em contraste com os métodos padrão nos quais os níveis de significância e confiança são baseados no modelo experimental completo (BIRNBAUM 1962, p.269).

2.6 POSTERIORI

Segundo Kruschke (2014), posteriori é uma componente do Teorema de Bayes usado para ponderar informações contidas em $p(\theta)$ e $f(\boldsymbol{x}|\theta)$ de maneira a obtermos $p(\theta|\boldsymbol{x})$:

$$p(\theta|\boldsymbol{x}) = \frac{p(\theta)L(\theta; \boldsymbol{x})}{\int p(\theta)L(\theta; \boldsymbol{x})d\theta} \propto p(\theta)L(\theta; \boldsymbol{x})$$

Onde:

- $p(\theta|\boldsymbol{x})$: A distribuição a posteriori do parâmetro θ dado os dados $\boldsymbol{x}$;
- $L(\theta; \boldsymbol{x})$: A verossimilhança dos dados $\boldsymbol{x}$ dado o parâmetro θ;

- $p(\theta)$: A distribuição a priori do parâmetro θ, que representa o conhecimento ou crença sobre o parâmetro antes de observar os dados;
- $p(\boldsymbol{x}) = \int p(\theta)L(\theta;\boldsymbol{x})d\theta$: Fator de normalização que assegura que a distribuição a posteriori seja uma distribuição de probabilidade válida.

Pode-se questionar, se, ao desprezar a constante de proporcionalidade não está sendo cometido algum equívoco formal. A resposta é que matematicamente sempre é possível recuperar essa constante mais tarde, se necessário. De acordo com Wasserman (2004, p.178), na prática, a constante de normalização $p(\boldsymbol{x})$ é frequentemente omitida durante a análise inicial, pois a forma da distribuição a posteriori é a mesma independentemente dessa constante. Isto é, estamos mais interessados na forma relativa da distribuição do que em sua magnitude absoluta.

Um ponto importante a considerar é a utilidade da distribuição a posteriori. Primeiramente, podemos usar essa distribuição para obter uma estimativa pontual do parâmetro que estamos investigando. Para isso, resumimos a informação central da distribuição a posteriori. Normalmente, utilizamos a média ou a moda dessa distribuição para obter essa estimativa (WASSERMAN, 2004).

A média a posteriori é frequentemente escolhida como a medida central, pois representa o valor esperado do parâmetro com base nos dados observados e na informação prévia.

2.7 AMOSTRAGEM DE MONTE CARLO

Segundo Gilks, Richardson e Spiegelhalter (1996), com problemas mais complexos e de alta dimensionalidade, métodos de amostragem como Monte Carlo via Cadeia de Markov (*Monte Carlo Markov Chain* - MCMC) podem ser utilizados para aproximar a integral de normalização. A amostragem de Monte Carlo é uma técnica estatística que utiliza números aleatórios para resolver problemas que podem ser determinísticos em princípio.

2.7.1.1 MCMC

Trazendo a termos leigos, dado que o próprio conceito de MCMC já seria digno de um trabalho por si só, podemos entender que uma cadeia de Markov é um modelo matemático que descreve uma sequência de eventos, onde a chance de cada evento depende apenas do evento anterior. Isso é chamado de propriedade de Markov, ou "propriedade sem memória". Basicamente, é como se a história completa do sistema pudesse ser resumida pelo seu estado atual, sem precisar saber como chegou lá.

Para entender melhor, imagine um jogo onde os jogadores se movem entre diferentes salas. A probabilidade de um jogador ir para a próxima sala depende apenas da sala em que ele está agora, e não das salas pelas quais já passou. As regras que determinam as chances do jogador se mover de uma sala para outra são as probabilidades de transição.

O método de Monte Carlo usando cadeias de Markov é uma técnica para resolver problemas complexos. Em vez de resolver o problema de uma só vez, ele transforma o problema em algo que pode ser resolvido passo a passo, através de uma série de movimentos. Cria-se uma simulação que vai de um estado a outro, como se estivesse caminhando entre as salas, até que essa simulação atinja uma situação estável, onde os estados param de mudar. Esta situação estável representa a solução do problema.

Em termos simples, é como se estivéssemos explorando um labirinto: nos movemos de uma sala para outra seguindo certas regras até que exploremos o suficiente para entender o layout completo do labirinto. O ponto em que paramos de aprender coisas novas sobre o labirinto é a solução do problema.

Entre os métodos MCMC, os mais citados são o amostrador de Gibbs e o algoritmo Metropolis-Hastings. Neste trabalho, o foco será o amostrador de Gibbs.

2.7.1.2 Amostrador de Gibbs

O amostrador de Gibbs, desenvolvido por Geman & Geman (1984) no contexto de restauração de imagens, é um método que utiliza um esquema markoviano de atualização. Ele possibilita a obtenção de amostras de uma distribuição conjunta através de iterações que envolvem amostragens das distribuições condicionais completas. Este método foi popularizado entre os bayesianos por Gelfand e Smith (1990), e desde então tem sido amplamente utilizado em diversas aplicações. Sua popularidade se deve à sua simplicidade conceitual e à facilidade de aplicação.

Seguindo Gelfand e Smith (1990) para descrever o amostrador de Gibbs, considere a densidade a posteriori conjunta $p(\boldsymbol{\theta}|\boldsymbol{x})$ onde $\boldsymbol{x}$ representa os dados e θ é o vetor dos parâmetros.

As distribuições de cada componente individual de θ podem ser especificadas, no mínimo proporcionalmente, a um determinado núcleo. Denotando as condicionais completas por:

$$p(\theta_i|x, \theta_1, \ldots, \theta_k), i = 1, \ldots, k$$

e o valor gerado para o i-ésimo parâmetro na j-ésima iteração por $\theta_i^{(j)}$, o algoritmo procede da seguinte forma:

I. Escolha de valores iniciais:

$$\theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_{k-1}^{(0)} \theta_k^{(0)}$$

II. Para j = 1, gere um valor $\theta_1^{(j)}$da distribuição condicional:

$$p\left(\theta_1 \middle| x, \theta_2^{(j-1)} \theta_3^{(j-1)}, \ldots, \theta_k^{(j-1)}\right)$$

gere um valor $\theta_2^{(j)}$ da distribuição condicional:

$$p\left(\theta_2 \middle| x, \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_k^{(j-1)}\right)$$

....

gere um valor $\theta_k^{(j)}$ da distribuição condicional

$$p\left(\theta_2 \middle| x, \theta_1^{(j)}, \theta_3^{(j)}, \ldots, \theta_{k-1}^{(j)}\right)$$

III. Repita o passo número ii. para j = 2, 3,..,m.

Sob condições gerais de regularidade, quando $m \rightarrow \infty$, os valores $(\theta_1^{(m)}, \dots, \theta_k^{(m)})$ convergem em distribuição para $p(\theta_1, \dots, \theta_k)$, ou seja, a distribuição da cadeia gerada pelo amostrador de Gibbs na iteração m converge em distribuição para a distribuição de equilíbrio, na norma de variação total, a uma taxa geométrica em m. Esta propriedade é conhecida como ergodicidade. Uma consequência importante deste resultado é que as médias ergódicas

$\bar{f}_m = \frac{1}{m}\sum_{j=1}^{m} f\,(\theta^{(j)})$ convergem quase certamente para $E[f(\theta)]$ quando $m \rightarrow \infty$, se a esperança calculada sob p, existir.

3 METODOLOGIA

3.1 MÉTODOS DE PESQUISA UTILIZADOS

Para o desenvolvimento deste trabalho, fez-se uma revisão bibliográfica como método de pesquisa principal. Esse método consistiu em uma análise de literatura acadêmica, incluindo artigos científicos, livros e outras publicações relevantes que abordam os conceitos fundamentais de inferência Bayesiana de parâmetros. A revisão bibliográfica permitiu a construção de uma base teórica sólida e atualizada, essencial para a compreensão e aplicação dos métodos estudados.

3.2 DESCRIÇÃO DO DESENVOLVIMENTO TEÓRICO

O desenvolvimento teórico deste trabalho concentrou-se em fundamentar os conceitos essenciais para a inferência Bayesiana de parâmetros. Onde abordou-se as definições e propriedades das distribuições a priori e a posteriori, destacando a importância das priori conjugadas e suas vantagens na simplificação das atualizações de conhecimento. Além disso, foi explorado a família exponencial de distribuições, que inclui muitas das distribuições de probabilidade mais comumente utilizadas em Estatística, e discutiu-se a existência de estatísticas suficientes de dimensão fixa. A teoria foi estruturada de maneira a fornecer um entendimento claro e coeso dos princípios que embasam os métodos Bayesianos.

3.3 DESCRIÇÃO DO DESENVOLVIMENTO PRÁTICO

3.3.1.Estudo I: Prioris Conjugadas

O primeiro estudo explora a teoria de prioris conjugadas para estimar a probabilidade p de obter "cara" ao lançar uma moeda. Suponhamos que p a priori siga uma distribuição $\text{Beta}(2, 2)$, que reflete uma crença inicial simétrica em torno de p= 0,5. Para realizar a simulação, lançaremos a moeda n= 100 vezes com uma verdadeira probabilidade p= 0,6 e observaremos o número de sucessos (caras), denotado por $p(\boldsymbol{x}|p) = p^t(1-p)^{n-t}$, onde $t = \sum_{i=1}^{n} x_i$ e $x_i = 0,1, i = 1, \ldots, n$. Utilizando a propriedade de conjugação das distribuições Beta e Binomial, a distribuição a posteriori de $(p|\boldsymbol{x})$ terá uma distribuição $Beta(2 + t, 2 + 100 - t)$. A distribuição a posteriori refletirá a atualização de nossa crença sobre p após observar os dados, ajustando-se de acordo com os sucessos observados e reduzindo a incerteza. Este estudo demonstrará como a inferência Bayesiana, utilizando prioris conjugadas, pode ser aplicada para atualizar nossas crenças sobre a probabilidade de um evento à medida que novas evidências são observadas.

3.3.2. Estudo II: Amostrador de Gibbs

O segundo estudo considera um exemplo clássico de um modelo Bayesiano com dois parâmetros, β_0 e β_1, e usa o amostrador de Gibbs para inferir suas distribuições a posteriori.

Consideramos um modelo de regressão simples da forma $yi = \beta_0 + \beta_1 xi + \epsilon_i$, onde $\epsilon_i \sim N(0, \sigma^2)$. Para simplificar, assumimos que σ^2=1. Definimos as distribuições a priori como $\beta_0 \sim \text{N}(0{,}10^2)$ e $\beta_1 \sim N(0{,}10^2)$. Utilizando o amostrador de Gibbs, obteremos amostras das distribuições a posteriori de β_0 e β_1.

Este estudo ilustrará a aplicação do amostrador de Gibbs para a inferência Bayesiana em um modelo de regressão.

4 RESULTADOS

4.1. ESTUDO I: PRIORIS CONJUGADAS

Foi realizado um experimento com o propósito de ilustrar o uso de prioris conjugadas no contexto da inferência bayesiana, particularmente destacando a relação entre as distribuições beta e binomial. Este estudo consistiu em estimar a probabilidade de obter "cara" ao lançar uma moeda, explorando as propriedades das prioris conjugadas e a atualização das distribuições a posteriori.

Iniciamos simulando 100 lançamentos de uma moeda, com uma probabilidade verdadeira de sucesso de p=0,6. Na etapa da inferência bayesiana, empregamos uma distribuição a priori $X \sim Beta(2, 2)$ para modelar nossa crença inicial sobre a probabilidade de sucesso. Após observar os dados, atualizamos a distribuição a posteriori utilizando o teorema de Bayes, resultando em uma distribuição $Beta(2 + t, 2 + 100 - t)$.

Este experimento teve como objetivo central demonstrar como as prioris conjugadas, como a distribuição beta para a probabilidade de sucesso em uma distribuição binomial, resultam em distribuições a posteriori que podem ser facilmente interpretadas e manipuladas. Esse aspecto fundamental da inferência bayesiana é de grande relevância na prática, pois simplifica o processo de atualização de crenças com base em novas evidências.

Tabela 1- Resultados obtidos a partir do Estudo de Simulação I (p=0,60)

Parâmetro	Estimativa	Intervalo Credibilidade	
		2,5%	97,5%
p	0,59615	0,50076	0,68805

Fonte: Os autores.

Observando os resultados apresentados na Tabela 1, nota-se a estimativa pontual de p é muito próxima do valor teórico esperado (0,60), e o intervalo de credibilidade obtido é relativamente estreito, o que sugere uma alta precisão na estimativa do parâmetro. A inclusão do valor teórico de 0,60 dentro do intervalo de credibilidade reforça a validade dos resultados obtidos.

4.2. ESTUDO II: AMOSTRADOR DE GIBBS

Primeiramente, foram simulados dados para uma regressão linear simples. A variável dependente é gerada a partir de uma combinação linear de uma variável independente, com coeficientes conhecidos β_0=2 e β_1=3 mais um termo de erro normal.

Foi gerada uma única cadeia via amostrador de Gibbs. Assim, não foi aplicado um método formal para diagnosticar a convergência; esta foi verificada informalmente pela proximidade dos valores estimados com os valores reais. Para cada iteração do algoritmo, os valores de β_0 e β_1 foram atualizados usando distribuições normais.

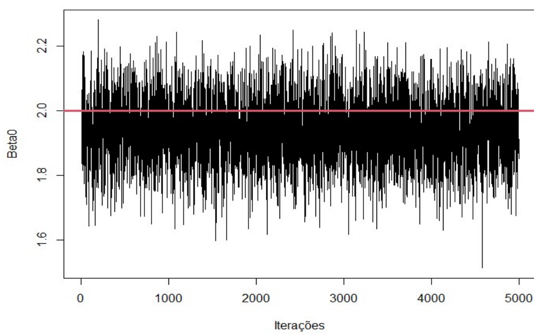
Após 5000 iterações, as medias dos valores amostrados de β_0 e β_1 foram usadas para calcular as estimativas pontuais e intervalos de credibilidade dos parâmetros.

Tabela 2- Resultados obtidos a partir do Estudo de Simulação II

Parâmetro	Estimativa	Intervalo Credibilidade	
		2,5%	97,5%
β_0	1,93902	1,73880	2,13760
β_1	2,98542	2,78596	3,17904

Fonte: Os autores.

Observando os resultados apresentados na Tabela 2, nota-se as estimativas pontuais de β0 e β1 são muito próximas do valores teóricos esperados (β0=2 e β1=3), e o intervalo de credibilidade obtido é relativamente estreito, sugeindo uma alta precisão na estimativa do parâmetro.



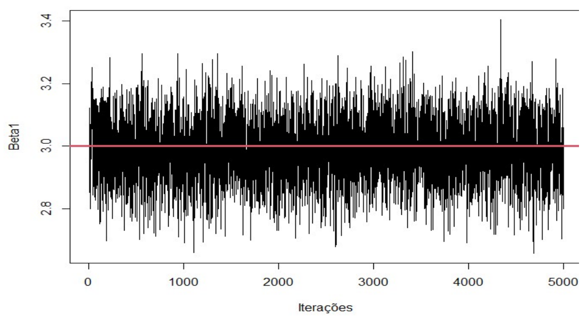


Figura 1: Trajetória para a cadeia gerada via amostrador de Gibbs, para os parâmetros β_0 e β_1.
Fonte: Os autores.

Por meio da análise da Figura 1, a trajetória da cadeia gerada mostra flutuações aleatórias em torno de valores constantes (sem tendências aparentes), isso indica informalmente que a cadeia aparente ter atingindo convergência.

Ademais, esse procedimento ilustrou como o amostrador de Gibbs pode ser empregado para realizar inferência bayesiana em um problema de regressão linear, permitindo não apenas estimar os parâmetros do modelo, mas também quantificar a incerteza associada a essas estimativas.

5 CONSIDERAÇÕES FINAIS

Este estudo buscou explorar a inferência bayesiana por meio da aplicação de duas técnicas fundamentais: simulação de distribuições conjugadas e o uso do amostrador de Gibbs em um modelo de regressão linear simples. Ao longo desta pesquisa, alguns aspectos relevantes foram identificados, proporcionando percepções valiosas para o entendimento dessas metodologias.

A metodologia adotada neste estudo permitiu uma compreensão aprofundada dos princípios fundamentais da inferência bayesiana. A simulação de distribuições conjugadas, como o caso da distribuição Beta conjugada para a distribuição binomial, proporcionou uma visão clara de como as distribuições a priori e a posteriori se relacionam, facilitando a análise e interpretação dos resultados.

Além disso, a implementação do amostrador de Gibbs para a estimação dos parâmetros em um modelo de regressão linear simples foi uma experiência enriquecedora. Esta técnica mostrou-se uma abordagem poderosa para a inferência bayesiana, permitindo não apenas a estimativa dos parâmetros do modelo, mas também a quantificação da incerteza associada a essas estimativas.

Durante a implementação das metodologias, foram identificadas algumas facilidades e dificuldades. Por um lado, a simulação de distribuições conjugadas apresentou-se relativamente simples de implementar, uma vez que as fórmulas analíticas para a distribuição a posteriori são conhecidas. Por outro lado, a implementação do amostrador de Gibbs exigiu um entendimento mais profundo dos conceitos estatísticos subjacentes, bem como habilidades de programação mais avançadas.

A literatura disponível corrobora consistentemente a eficácia das metodologias exploradas neste trabalho, destacando tanto suas vantagens quanto suas limitações. A inferência bayesiana é especialmente valiosa em cenários onde a disponibilidade de dados é limitada ou a estrutura do modelo é complexa.

Uma das principais forças dessa abordagem é sua capacidade de incorporar conhecimento prévio através da distribuição a priori, o que pode melhorar a robustez das estimativas quando os dados são escassos ou incompletos. Além disso, a inferência bayesiana oferece uma forma natural de quantificar a incerteza das estimativas, fornecendo distribuições completas para os parâmetros ao invés de simples estimativas pontuais.

No entanto, a inferência bayesiana também apresenta desafios e limitações. A especificação das distribuições a priori pode ser subjetiva e influenciar significativamente os resultados, especialmente em situações onde há pouca ou nenhuma informação prévia confiável. Além disso, os métodos MCMC, como o amostrador de Gibbs, podem ser computacionalmente intensivos, especialmente para modelos de alta dimensionalidade ou quando as distribuições a posteriori são complexas e difíceis de amostrar. Em tais casos, a convergência das cadeias MCMC pode ser lenta, exigindo um número muito grande de iterações para obter amostras representativas da distribuição posterior.

Propostas futuras incluem a exploração de técnicas mais avançadas de inferência bayesiana, como o algoritmo de Metropolis-Hastings e o Monte Carlo Hamiltoniano, que também fazem parte da família MCMC. O algoritmo de Metropolis-Hastings é uma técnica generalizada que pode ser aplicada a uma ampla gama de distribuições, ajustando novos valores baseados em uma distribuição proposta e aceitando ou rejeitando esses valores conforme uma probabilidade específica. Este algoritmo é particularmente útil em situações onde as distribuições a posteriori são complexas e não se encaixam bem em distribuições conjugadas. Por outro lado, o Monte Carlo Hamiltoniano (ou Hamiltonian Monte Carlo - HMC) utiliza a física hamiltoniana para explorar o espaço das amostras de maneira mais eficiente, especialmente em distribuições de alta dimensionalidade. O HMC é indicado para modelos com muitas variáveis, onde a alta dimensionalidade torna os métodos tradicionais de MCMC ineficazes, pois ele pode reduzir a correlação entre amostras sucessivas e acelerar a convergência.

Em suma, este estudo proporcionou uma sólida compreensão das metodologias de inferência bayesiana, demonstrando seu potencial e relevância em uma variedade de aplicações estatísticas e científicas. A contínua exploração e desenvolvimento dessas técnicas prometem contribuir significativamente para o avanço do conhecimento, especialmente em áreas como machine learning, onde a capacidade de incorporar conhecimento prévio e quantificar a incerteza das estimativas traz vantagens significativas. Essas abordagens estão beneficiando campos como processamento de linguagem natural, visão computacional e modelagem preditiva, oferecendo soluções mais robustas e adaptativas para problemas complexos.

REFERÊNCIAS

BATISTA, D. Inferência Bayesiana. Disponível em: <https://bendeivide.github.io/handouts/infbayes/>. Acesso em: 19 jun. 2024.

BERNARDO, José M.; SMITH, Adrian F. M. Bayesian Theory. Chichester: Wiley, 2000.

BERGER, J. O. (1985). Statistical Decision Theory and Bayesian Analysis (2nd ed.). Springer-Verlag.

BIRNBAUM, Allan. On the Foundations of Statistical Inference. Berkeley: University of California Press, 1962.

CASSELLA, George; BERGER, Roger L. *Statistical Inference*. 2. ed. Belmont: Duxbury, 2001. (GELMAN et al., 2013).

EHLERS, R.S. Introdução a Inferência Bayesiana. Disponível em < http://www.leg.ufpr.br/~paulojus/CE227/ce227/ce227.html> . Acesso em: 19 jun. 2024.

GELFAND, A.E., SMITH, A.F.M. Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc., 1990.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A., RUBIN, D. B. ***Bayesian Data Analysis*** (3rd ed.). CRC Press, 2013.

GILKS, W. R., RICHARDSON, S., SPIEGELHALTER, D. J. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC, 1996.

JEFFREYS, H. Theory of Probability (3rd ed.). Oxford, UK: Oxford University Press, 1961.

KRUSCHKE, J. K. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd ed. Academic Press, 2014.

LUNN, D.; JACKSON, C.; BEST, N.; THOMAS, A.; SPIEGELHALTER, D. The BUGS Book: A Practical Introduction to Bayesian Analysis. CRC Press, 2012.

ROBERT, C. P., CASELLA, G. Monte Carlo Statistical Methods (2nd ed.). Springer, 2014.

PRESS, S. James. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. 2. ed. New York: Wiley, 2003.

ZABALA, F. Inferência Bayesiana. Disponível em: <https://filipezabala.com/eb/inferencia-bayesiana.html.> Acesso em: 19 jun. 2024.

WASSERMAN, L. All of Statistics: *A Concise Course in Statistical Inference*. New York: Springer, 2004. p. 178. Disponível em:<https://egrcc.github.io/docs/math/all-of-statistics.pdf.> Acesso em: 19 jun. 2024.

APÊNDICE A – Códigos em Linguagem R dos Estudos de Simulação

```r
##########################ESTUDO I: PRIORIS CONJUGADAS##############################
library(gridExtra)
library(ggplot2)

# Simulação de dados
set.seed(123)
n <- 100
p_true <- 0.6
data <- rbinom(n, size = 1, prob = p_true)
k <- sum(data)

# Inferência Bayesiana
alpha_prior <- 2
beta_prior <- 2
alpha_post <- alpha_prior + k
beta_post <- beta_prior + n - k

p_bayes <- alpha_post / (alpha_post + beta_post)

# Intervalo de Credibilidade para p
ci_bayes_lower <- qbeta(0.025, alpha_post, beta_post)
ci_bayes_upper <- qbeta(0.975, alpha_post, beta_post)

cat("Estimativa Bayesiana de p:", p_bayes, "\n")
cat("Intervalo de Credibilidade de 95% para p: [", ci_bayes_lower, ", ", ci_bayes_upper, "]\n")

```

```r
################################ESTUDO II: AMOSTRADOR DE GIBBS############################################

# Simulando dados
set.seed(123)
n <- 100
x <- rnorm(n, 0, 1)
beta0_true <- 2
beta1_true <- 3
y <- beta0_true + beta1_true * x + rnorm(n, 0, 1)

# Ajuste do modelo de regressão linear
model <- lm(y ~ x)
summary(model)


# Inferência Bayesiana com amostrador de Gibbs
gibbs_sampler <- function(y, x, n_iter = 1000, mu0 = 0, tau0 = 10) {
  n <- length(y)
  beta0_samples <- numeric(n_iter)
  beta1_samples <- numeric(n_iter)

  beta0 <- 0
  beta1 <- 0

  for (i in 1:n_iter) {
    y_star <- y - beta1 * x
    var_beta0 <- 1 / (n + 1/tau0^2)
    mean_beta0 <- var_beta0 * sum(y_star)
    beta0 <- rnorm(1, mean_beta0, sqrt(var_beta0))

    x_star <- x
    y_star <- y - beta0
    var_beta1 <- 1 / (sum(x_star^2) + 1/tau0^2)
    mean_beta1 <- var_beta1 * sum(x_star * y_star)
    beta1 <- rnorm(1, mean_beta1, sqrt(var_beta1))

    beta0_samples[i] <- beta0
    beta1_samples[i] <- beta1
  }

  list(beta0 = beta0_samples, beta1 = beta1_samples)
}

samples <- gibbs_sampler(y, x, n_iter = 5000)

beta0_estimate_bayes <- mean(samples$beta0)
beta1_estimate_bayes <- mean(samples$beta1)

plot(samples$beta0, type="l", xlab="Iterações", ylab = "Beta0")
abline(h=beta0_true, lwd=3, col=2)

plot(samples$beta1, type="l", xlab="Iterações", ylab = "Beta1")
abline(h=beta1_true, lwd=3, col=2)


# Intervalo de Credibilidade para beta0 e beta1
ci_bayes_beta0 <- quantile(samples$beta0, c(0.025, 0.975))
ci_bayes_beta1 <- quantile(samples$beta1, c(0.025, 0.975))

cat("Estimativa Bayesiana de beta0:", beta0_estimate_bayes, "\n")
cat("Estimativa Bayesiana de beta1:", beta1_estimate_bayes, "\n")
cat("Intervalo de Credibilidade de 95% para beta0: [", ci_bayes_beta0[1], ", ", ci_bayes_beta0[2], "]\n")
cat("Intervalo de Credibilidade de 95% para beta1: [", ci_bayes_beta1[1], ", ", ci_bayes_beta1[2], "]\n")
```