CE-315: Estatística Bayesiana

Banco de Questões

2026-10-12

Este documento reúne questões de estudo da disciplina adicionais aos materiais disponibilizados. As questões são organizadas por tópico. Novas questões são disponibilizadas ao longo do curso, à medida que os assuntos correspondentes são vistos em aula.

Em um grupo de estudantes, 55% são do curso A, 35% do curso B e o restante do curso C. A proporção de estudantes que estão fazendo a segunda graduação em cada curso é de 5, 10 e 25%, respectivamente. Um estudante é sorteado ao acaso e verifica-se que está fazendo a segunda graduação.

a. A qual curso você arriscaria dizer que o estudante sorteado pertence, sem fazer contas? Justifique.
b. Qual a probabilidade de ser de cada um dos cursos? Resolva o problema da forma que achar adequada.
c. Considere agora o problema do ponto de vista Bayesiano. Qual é a variável resposta (Y) e sua distribuição? Qual é o parâmetro (θ) e sua distribuição a *priori*? Qual a verossimilhança? E a *posteriori*? E a marginal (preditiva)?
d. Esboce um gráfico da *priori* e da *posteriori*.
e. Interprete os resultados.
f. Agora que resolveu o problema, em qual curso "apostaria" que o estudante pertence? Justifique.

Uma empresa adquire 30% de sua matéria prima de um primeiro fornecedor, 50% de um segundo e o restante de um terceiro fornecedor. Os lotes de matéria prima são inspecionados e, se considerados não satisfatórios, são enviados de volta. Sabe-se que 2%, 5% e 1% dos lotes são retornados a cada fornecedor, respectivamente. Rejeitando-se um lote, deseja-se quantificar a chance de ter vindo de cada um dos fornecedores.

a. Resolva o problema da forma que achar adequada.
b. Considere agora o problema do ponto de vista Bayesiano. Qual é a variável resposta (Y) e sua distribuição? Qual é o parâmetro (θ) e sua distribuição a *priori*? Qual a verossimilhança? E a *posteriori*?
c. Esboce um gráfico da *priori* e da *posteriori*.

Os dados deste problema são hipotéticos, sem nenhuma base real. Suponha que os maiores partidos da Inglaterra, "Labour", "Conservative", "LibDems", obtiveram 35, 42 e 20% dos votos em uma eleição recente. O restante foi distribuído entre partidos menores. Tem-se ainda que os percentuais de favoráveis ao BREXIT em cada um destes partidos é de 15, 62 e 40% respectivamente. Não se sabe sobre os demais mas assume-se que seja de 50%. Entrevistou-se uma pessoa que se declarou favorável ao BREXIT.

a. Qual a probabilidade de que ela seja apoiadora do "Labour"? E dos "LibDem"?
b. Escreva os dados e quantidades de interesse na notação geral de variável observada Y e parâmetro θ.
c. Forneça a distribuição a priori do problema.
d. Forneça a verossimilhança do problema.
e. Encontre a distribuição a posteriori.
f. Esboce um gráfico com as distribuições priori e posteriori (e a preditiva, se pedido).
g. E se a pessoa entrevistada se declarasse **contra** o Brexit — como ficam a verossimilhança e a posteriori?

(Migon, Gamerman & Louzada, 2015) Um teste para verificar se um motorista está dirigindo bêbado tem uma chance de 0,8 de estar correto, isto é, de fornecer resultado positivo quando o motorista está com níveis de álcool no sangue acima do permitido ou negativo quando está abaixo do limite aceitável. Um segundo teste (muito mais caro) é aplicado apenas aos casos suspeitos. Ele nunca falha se o motorista não está irregular, mas tem uma chance de erro de 10% se o motorista está acima do limite. Supondo que 25% dos motoristas parados estão acima do limite, calcule:

a. a proporção de motoristas parados que serão submetidos ao segundo teste;
b. a probabilidade a posteriori de que um motorista realmente esteja com níveis de álcool acima do limite, se foi submetido e positivou em ambos os testes;
c. a proporção de motoristas que serão submetidos apenas ao primeiro teste;
d. monte uma notação adequada ao problema indicando as quantidades de interesse no contexto bayesiano.

Dentre usuários que se cadastram em uma plataforma, 20% declaram ter menos de 25 anos e, destes, acredita-se que 10% utiliza perfil falso. Entre 25 e 50 anos são 50% dos usuários, com 20% deles sendo perfis falsos. Acima de 50 anos são 30% dos usuários e 5% deles com perfis falsos. Toma-se um usuário ao acaso e verifica-se que possui perfil falso.

a. Sem fazer contas, qual faixa etária você espera ser a deste usuário? Justifique.
b. Qual a probabilidade de ser de cada uma das faixas etárias? Resolva o problema da forma que achar adequada.
c. Considere agora o problema do ponto de vista Bayesiano: qual é a variável resposta (Y) e sua distribuição? Qual é o parâmetro (θ) e sua distribuição a *priori*? Qual a verossimilhança? E a *posteriori*? E a marginal (preditiva)?
d. Esboce um gráfico da *priori* e da *posteriori*.
e. Interprete os resultados. A qual faixa etária você espera que o usuário pertença?
f. As proporções de perfis masculinos são de 15%, 30% e 55% para as faixas etárias de menos de 25, entre 25 e 50 e acima de 50 anos, respectivamente. Se o usuário de perfil falso selecionado for do sexo masculino, qual a probabilidade de pertencer a cada uma das faixas etárias?

Dados obtidos de um mercado mostram que três corretoras A, B e C são responsáveis por 20%, 50% e 30% dos contratos negociados, respectivamente. Do volume de cada corretora, 20%, 5% e 2%, respectivamente, são contratos futuros em dólares. Em uma auditoria, toma-se um contrato ao acaso, sem identificação da corretora, e verifica-se que é futuro em dólares.

a. Qual a probabilidade de ter sido negociado por cada uma das corretoras?
b. Considere agora o problema do ponto de vista Bayesiano: variável resposta, parâmetro e priori, verossimilhança, posteriori e marginal (preditiva).

c. Esboce em um gráfico a *priori* e a *posteriori*.
d. Interprete os resultados.

Apurou-se dados adicionais que mostram que, do volume de cada corretora, os contratos de pessoa física são 80%, 90% e 60%, respectivamente.

e. Se um contrato é tomado ao acaso e apenas se sabe que é de pessoa jurídica, qual a probabilidade de que seja de cada uma das corretoras?
f. Se um contrato é tomado ao acaso e sabe-se que é de pessoa física e não é futuro em dólar, de qual corretora é mais provável ser o contrato?

Considere o algoritmo *Naïve Bayes*.

a. Explique o algorítmo para classificação, utilizando o contexto da questão de perfis falsos por faixa etária. Quais são as características, vantagens e desvantagens deste método?
b. Crie um exemplo hipotético com uma variável de classificação com três níveis e três classificadoras. Explique como o algoritmo naïve Bayes poderia ser aplicado, descrevendo os passos do algoritmo.

---

a. Escreva de maneira sucinta (um parágrafo com no máximo 10 linhas) como você explicaria para um pesquisador de outra área as características da análise bayesiana em contraste com outra(s) abordagem(ns) para inferência estatística.
b. Usando apenas palavras, como você explicaria para um leigo o que é inferência bayesiana?
c. Usando notação estatística, como você explicaria para um estatístico (que ainda não sabe) o que é inferência bayesiana?
d. Quais são as principais diferenças entre a inferência clássica (frequentista e por verossimilhança) e a inferência bayesiana? (Por exemplo, para cada uma: qual é o objeto de inferência? Quais as suposições feitas sobre os parâmetros? Como os dados são tratados? Como as informações prévias podem ser consideradas?)
e. O que você julga serem os pontos fortes e fracos da inferência bayesiana?
f. Explique distribuição amostral, função de verossimilhança e distribuição *a posteriori* e como estão relacionadas a *paradigmas* de inferência estatística.

Forneça a expressão da priori conjugada dos seguintes modelos de $[Y|\theta]$. Justifique as respostas.

a. Geométrica(θ)
b. Gama($a = 5, \theta$)
c. Normal($\mu = 100, \theta$)
d. Weibull($\alpha = 2, \theta$)
e. Normal($\theta, \sigma = 1$)

Embora prioris conjugadas tornem a computação bayesiana extremamente simples, elas podem não ser apropriadas e por vezes simplesmente não existem (de forma útil) para o modelo que se deseje ajustar.

a. Explique como análises podem ser conduzidas neste caso, considerando pelo menos duas abordagens.
b. Implemente as abordagens mencionadas em um problema de sua escolha para distribuições da variável e do parâmetro.

Mostre como as abordagens mencionadas no item anterior seriam aplicadas no exemplo a seguir. Inclua nos comentários como resumos da posteriori tais como média, variância, quantis e intervalos seriam obtidos

em cada abordagem.

Exemplo (aproveitamento de saques em jogos de tênis). Considere dados (iid) $x = (x_1, \ldots, x_n)$ das taxas de sucesso no primeiro saque de um jogador de tênis em n jogos de um campeonato. Assuma o modelo

$$X|\theta \sim f(x_i|\theta) = \theta(\theta+1)\, x_i^{\theta-1}(1-x_i), \qquad x_i \in (0,1),\ \theta > 0.$$

Não existe uma família conjugada usual para este modelo e considera-se uma priori gama $\theta \sim \mathrm{G}(a,b)$.

Considere agora que $n = 20$, $\sum_i \log(x_i) = -4{,}59$ e suponha que a priori é definida com $a = b = 1$. Mostre como estes dados seriam utilizados na obtenção das expressões relevantes das abordagens mencionadas. Discuta e explique procedimentos e passos adotados em situações em que não há expressões analíticas (fechadas).

Explique os tópicos a seguir, incluindo exemplos, tendo como referência as aulas e materiais do curso:

a. Priori imprópria.
b. Princípio da verossimilhança.
c. Reparametrizações, prioris equivalentes e invariância.

Seja $y_1, \ldots, y_n$ uma amostra aleatória de uma distribuição normal de média μ desconhecida e variância σ^2 conhecida, com priori $\mu \sim \mathrm{N}(m_0, \tau_0^2)$.

a. Obtenha a expressão da verossimilhança em função de μ.
b. Mostre que a posteriori de μ também é normal e obtenha seus parâmetros.
c. Escreva a média a posteriori como uma média ponderada entre m_0 e $\bar{y}$, identificando os pesos em termos das precisões $1/\tau_0^2$ e n/σ^2. O que ocorre com o peso da priori quando $n \to \infty$?
d. Obtenha a distribuição preditiva para uma nova observação y_{n+1}.

Seja $x_1, \ldots, x_n$ uma a.a. de uma distribuição normal de média μ conhecida e variância σ^2 desconhecida.

a. Obtenha a expressão da verossimilhança do modelo.
b. Obtenha a priori de Jeffreys para σ^2.
c. Obtenha a expressão da distribuição a posteriori.
d. É possível identificar a posteriori do modelo como alguma distribuição conhecida? Qual?
e. Considere que foi tomada a amostra a seguir, com $\mu = 10$. Obtenha a expressão da posteriori.

$$12{,}1\ ;\ 8{,}7\ ;\ 11{,}3\ ;\ 9{,}2\ ;\ 10{,}5\ ;\ 9{,}7\ ;\ 11{,}6$$

f. Indique (com comandos do R ou de alguma outra forma) como resumos pontuais e intervalares desta distribuição a posteriori poderiam ser obtidos para fins de inferência.

Seja Y uma variável aleatória com distribuição normal de média igual a zero e precisão τ, sendo $\tau = 1/\sigma^2$. Será tomada uma amostra de n observações $y_1, \ldots, y_n$ deste modelo.

a. Qual a priori conjugada para τ?
b. Qual é a distribuição posteriori para a priori do item anterior?
c. Qual a priori de Jeffreys para τ?
d. Qual é a distribuição posteriori para a priori de Jeffreys?

Suponha que o tempo em minutos necessário para atender um cliente em um banco possui uma distribuição exponencial com parâmetro θ. Para fins de planejamento de serviços deseja-se conhecer melhor o processo, o que é feito por inferências sobre θ. Baseada em conhecimento prático e subjetivo, a incerteza sobre θ é descrita por uma distribuição *a priori conjugada*.

a. Escreva a função de probabilidades/densidade para as observações.
b. Escreva a função de verossimilhança para uma amostra $(y_1, \ldots, y_n)$.
c. Reconheça uma distribuição de probabilidades proporcional a esta expressão da verossimilhança. Qual é a distribuição e com quais parâmetros?
d. Forneça a expressão de uma priori conjugada. Acredita-se que a incerteza sobre θ é descrita a priori por uma distribuição Gama com média 0,25 e desvio padrão 1. Encontre os hiperparâmetros da priori.
e. Obtenha a expressão da distribuição marginal (preditiva a priori). Identifique que a distribuição é uma Pareto (Lomax), conforme o formulário do curso.
f. Obtenha o valor esperado da distribuição marginal, e descreva formas de simular valores dela.
g. Obtenha a expressão da posteriori e indique a distribuição de probabilidades. Quais estatísticas da amostra são necessárias para a verossimilhança e a posteriori?
h. Se o tempo médio observado para atender 20 clientes é de 10,5 minutos, determine a distribuição a posteriori para θ.
i. Qual o valor esperado da posteriori? Mostre que ele é uma média ponderada entre o valor esperado da priori e o estimador de máxima verossimilhança.
j. Qual seria a distribuição preditiva para uma nova observação, após obter esta primeira amostra? Qual o valor esperado desta marginal (descreva ao menos três formas de obtê-lo)?
k. Em um segundo período de observação, obteve-se uma nova amostra (tamanho e média diferentes). Qual a posteriori após incorporar também esses dados? Que suposição é feita ao combinar as duas amostras desta forma?
l. Forneça os comandos para desenhar em um gráfico as curvas da priori, verossimilhança e posteriori.

Seja $y_1, \ldots, y_n$ uma amostra aleatória de uma distribuição com função de probabilidades $f(y|\theta) = \theta \exp\{-\theta y\}$.

a. Obtenha a priori de Jeffreys para θ.
b. Obtenha a expressão da posteriori e da distribuição preditiva (marginal a priori e a posteriori).

Seja o contexto da exponencial acima, mas deseja-se usar como priori uma distribuição Log-Normal (portanto não conjugada), com posteriori que não é uma distribuição "padrão".

a. Explique três formas de se obter uma aproximação para a posteriori.
b. Explique como seriam obtidas simulações da preditiva a priori.
c. Explique como seriam obtidas simulações da preditiva a posteriori.

Nesta família de questões um experimento é repetido de forma independente até se obter o r-ésimo resultado positivo (amostragem inversa / binomial negativa), e deseja-se fazer inferência sobre a probabilidade θ de sucesso.

Parte 1 — imagens médicas. Decidiu-se examinar um conjunto de imagens médicas, tomadas independentemente, para avaliar a proporção com determinada característica morfológica. Em um primeiro estudo tomaram-se 9 imagens, sendo detectadas 3 com a característica; em um segundo estudo decidiu-se

amostrar até obter 3 com a característica, ocorrendo que também foram necessárias 9 no total (mesmos dados, dois desenhos amostrais diferentes).

a. Como seria testada a hipótese de interesse em cada estudo no enfoque não bayesiano? Os dois estudos forneceriam a mesma conclusão estatística?
b. Em uma análise Bayesiana com priori $\theta \sim$ Be(1,5; 1,5), qual seria a posteriori em cada estudo? Como deveria ser avaliada a posteriori para concluir sobre a hipótese de interesse? Os estudos produziriam a mesma conclusão? Justifique.
c. Se alternativamente decide-se adotar a priori de Jeffreys em cada estudo, elas seriam as mesmas e produziriam a mesma posteriori? Justifique.

Parte 2 — processo seletivo. Modela-se com a binomial negativa o número de entrevistas necessárias para se conseguir três candidatos aprovados em uma fase inicial de um processo seletivo. Não se tem ideia da probabilidade de aprovação, e decide-se por uma priori que reflita essa ausência de informação (vaga/de Jeffreys). Em uma rodada de seleção foram necessárias 20 entrevistas para se conseguir os três aprovados.

d. Escreva a função de densidade de probabilidades para a probabilidade de aprovação antes da rodada de seleção.
e. Escreva a função de verossimilhança.
f. Reconheça uma distribuição de probabilidades proporcional a esta expressão da verossimilhança. Qual é a distribuição e com quais valores de parâmetros?
g. Forneça a expressão de uma distribuição a priori conjugada que poderia ser utilizada caso houvesse alguma suspeita sobre valores não uniformes para a probabilidade de aprovação.
h. Qual(ais) valor(es) você especificaria para o(s) parâmetro(s) da distribuição priori conjugada (hiperparâmetros) neste caso? Justifique.
i. Como é obtida a distribuição marginal (preditiva a priori)?
j. Explique como faria para simular valores desta distribuição marginal.
k. Obtenha a expressão da posteriori e indique qual a distribuição de probabilidades.
l. Qual(ais) quantidade(s) calculada(s) com a amostra são necessárias para a obtenção da verossimilhança e posteriori?
m. Forneça a expressão da distribuição a posteriori para θ com os dados da rodada de seleção (20 entrevistas, 3 aprovados).
n. Qual o valor esperado da distribuição a posteriori para θ?
o. Como poderia ser obtida a distribuição preditiva para uma nova observação após obter esta primeira amostra?
p. Como obter o valor esperado da preditiva do item anterior?
q. Foi necessário fazer uma segunda rodada de seleção na qual foram necessárias 15 entrevistas para se conseguir os 3 aprovados. Qual a posteriori após também considerar os dados desta segunda rodada de seleção?

Parte 3 — reações químicas. Um experimento envolvendo uma série de reações químicas, que resulta ou não no produto desejado, foi repetido consecutivamente (de forma independente) até que o terceiro resultado positivo fosse obtido, sendo necessária a realização de 12 experimentos no total.

r. Encontre a priori de Jeffreys para θ diretamente a partir do modelo Binomial Negativo e obtenha a posteriori correspondente.

Seja $Y|\theta \sim \text{Bin}(n, \theta)$ com priori de θ dada por uma **mistura** das distribuições Beta(2,10) e Beta(4,2), com pesos 0,20 e 0,80, respectivamente. Uma amostra forneceu $y = 6, n = 20$.

a. Esboce os gráficos de cada um dos componentes da priori.
b. Obtenha a expressão da posteriori e um esboço do seu gráfico.

Seja $y_1, \ldots, y_n$ uma amostra aleatória de uma distribuição com função de probabilidades

$$f(y|\theta) = \frac{\mathrm{e}^{-\theta}\theta^y}{y!}\,; \quad y = 0, 1, \ldots$$

e assume-se como distribuição a priori

$$f(\theta) = \mathrm{e}^{-\theta}\,; \quad 0 < \theta.$$

Obtenha a expressão da posteriori e da distribuição preditiva.

O número de defeitos em um rolo de 1200 metros de uma fita magnética possui distribuição Poisson(θ). A distribuição priori para θ é Gama(3; 1). Quando cinco rolos deste tipo são selecionados ao acaso, o número de defeitos encontrados em cada um deles é: 2, 2, 6, 0 e 3, respectivamente. Determine a distribuição a posteriori de θ.

A distribuição de falhas ao longo do comprimento de uma fibra artificial segue um processo de Poisson, de modo que o número de falhas em um comprimento l da fibra é Po($l\theta$). Pouco se sabe sobre θ. O número de falhas obtido em 5 fibras de comprimentos 10, 15, 25, 30 e 40 metros, respectivamente, foi de 3, 2, 7, 6 e 10. Encontre a distribuição preditiva para o número de falhas em outra fibra com 60 metros de comprimento. Explique os passos que seriam executados para obter amostras e resumos da preditiva por simulação estocástica.

Foi visto no curso que para o modelo (verossimilhança) Poisson a priori conjugada é uma distribuição Gama. Entretanto pode-se pensar em outras prioris, tal como uma log-normal, ou seja, um modelo com $Y \sim \mathrm{P}(\lambda)$ e $\log(\lambda) \sim \mathrm{N}(a, b)$.

a. É possível obter uma posteriori em forma fechada e conhecida neste caso? Qual a expressão da posteriori (a menos de uma constante de proporcionalidade, se necessário)?
b. Se você tivesse dados de contagem para analisar sob o paradigma Bayesiano com este modelo, como procederia? Qual(is) procedimento(s) adotaria para obter inferências a posteriori?

Vai se modelar o número diário de acidentes em um trecho de rodovia com o modelo Poisson com priori "half-normal".

a. Mostre a expressão da posteriori (a uma constante) e indique se é (ou não) compatível com alguma distribuição conhecida.
b. Opta-se por fazer uma análise com aproximação discreta de cinco pontos. O conjunto de dados é $0, 3, 2, 1, 0, 2$. Escolha cinco pontos razoáveis e obtenha a posteriori (discreta) e calcule sua média e moda.
c. Implemente computacionalmente a solução: aproximação de 5 pontos e depois discretização com mais pontos; escreva uma função para a posteriori que inclua o cálculo da constante normalizadora por integração numérica.

O número de atendimentos externos que um chaveiro executa por dia é modelado por uma distribuição de Poisson com parâmetro θ. Baseado em conhecimento prático e subjetivo, acredita-se que a incerteza sobre θ é descrita a priori por uma distribuição Gama com média 3,2 e desvio padrão 0,4.

a. Declare o modelo bayesiano completo, escrevendo as expressões das distribuições envolvidas.
b. Encontre os hiperparâmetros da priori conjugada a partir da média e do desvio padrão dados.
c. Um estudo de 10 dias registrou um total de 16 atendimentos. Obtenha a distribuição a posteriori de θ e seu valor esperado.
d. Obtenha a distribuição preditiva para o número de atendimentos em um novo dia.

Defina a priori de Jeffreys e comente sobre seu potencial de uso, características importantes e possível(eis) limitação(ões).

Encontre a priori de Jeffreys para θ no modelo geométrico

$$f(x|\theta) = (1-\theta)^{x-1}\theta\,; \quad x = 1, 2, ...$$

(em uma variante, a parametrização é $f(y|\theta) = (1-\theta)^{y}\theta$, $y = 0, 1, 2, ...$) para uma amostra $x_1, \ldots, x_n$.

a. Obtenha as expressões da posteriori e da preditiva.
b. Considere a amostra $5, 4, 7, 3, 2, 5, 6, 8, 10$. Forneça a expressão da posteriori e um valor predito para uma nova observação.
c. Escolha e fixe um valor razoável para o parâmetro e simule uma amostra de tamanho $n = 10$. Explore graficamente o formato da priori. Mostre que a priori é imprópria. Obtenha os gráficos da posteriori e da preditiva posteriori.

Suponha que Y tenha uma distribuição de Pareto $\text{Pa}(a, \theta)$, em que a é conhecido mas θ é desconhecido:

$$f(y|\theta) = \theta\, a^{\theta}\, y^{-\theta-1}\,; \quad y > a, \qquad \text{E}[Y] = \frac{a\theta}{\theta - 1} \text{ para } \theta > 1.$$

a. Encontre a priori de Jeffreys e a correspondente distribuição posteriori para θ.
b. Explique como poderiam ser obtidos resumos (média, mediana, quantis, intervalos) da distribuição posteriori neste caso.

Considere o contexto de predição (a posteriori) Bayesiana, porém deseja-se obter somente o valor esperado e a variância da preditiva. Obtenha as expressões destas quantidades no modelo de verossimilhança Poisson com priori Gama.

Explique de forma concisa a diferença entre intervalos frequentista, de verossimilhança e bayesiano (de quantis ou HDI).
