Inferência Bayesiana

1. Modela-se com a distribuição binomial negativa o número de entrevistas necessárias para se conseguir três candidatos aprovados em uma fase inicial de um processo seletivo. Considera-se uma abordagem Bayesiana. Não se tem ideia da probabilidade de aprovação e decide-se por uma distribuição *a priori* que reflita tal situação. Em uma rodada de seleção foram necessárias 20 entrevistas para se conseguir os três aprovados.

   (a) Escreva a função de densidade de probabilidades para a probabilidade de aprovação antes da rodada de seleção.

   (b) Escreva a função de verossimilhança.

   (c) Reconheça uma distribuição de probabilidades proporcional à esta expressão da verossimilhança. Qual é a distribuição e com quais valores de parâmetros?

   (d) Forneça a expressão de uma distribuição à *priori* conjugada que poderia ser utilizada caso houvesse alguma suspeita sobre valores não uniformes para a probabilidade de aprovação.

   (e) Qual(ais) valor(es) você especificaria para o(s) parâmetro(s) da distribuição *priori* conjugada (hiper-parâmetros) neste caso? Justifique.

   (f) Como é obtida a distribuição marginal (preditiva a priori)?

   (g) Explique como faria para simular valores desta distribuição marginal.

   (h) Obtenha a expressão da posteriori e indique qual a distribuição de probabilidades.

   (i) Qual(ais) quantidade(s) calculada(s) com a amostra são necessárias para a obtenção da verossimilhança e posteriori?

   (j) Forneça a expressão da distribuição a posteriori para θ com os dados da rodada de seleção.

   (k) Qual o valor esperado da distribuição a posteriori para θ?

   (l) Como poderia ser obtida a distribuição preditiva para uma nova observação após obter esta primeira amostra?

   (m) Como obter o valor esperado da preditiva do item anterior?

   (n) 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?

2. Dentre usuários que se cadastram em uma plataforma, 20% declaram ter menos de 25 anos e acredita-se que destes, 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 falso. Toma-se um usuário ao acaso e verifica-se que possui perfil falso.

   (a) Sem fazer contas, qual faixa etária que voce espera ser a deste usuário? Justifique.

   (b) Qual a probabilidade de ser de cada uma das faixas etárias? Resolva a 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?0
      - 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, a qual faixa etária voce espera o usuário pertencer? Justifique.

   (g) As proporções de perfis masculinos são de 15%, 30% e 55% para as faixas etárias de menos de 25 anos, entre 25 e 50 anos 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?

3. Indique para cada trecho de código a seguir o modelo usado e o que está sendo obtido.

   (a)
   ```
   vec1 <- runif(1E5)
   vec2 <- rbinom(1E5, size = 20, prob = vec1)
   plot(table(factor(vec2, levels = 0:20)), type = 'h')
   mean(vec2 > 5)
   vec3 <- rbeta(1E5, 8+1, (20-8)+1)
   hist(vec3, freq = FALSE, nclass = 30)
   mean(vec3 < 0.30)
   ```

   (b)
   ```
   curve(dbeta(x, 2, 8), ylim = c(0, 7))
   curve(dbeta(x, 2 + 10, 8 + 30), col = 4, add = TRUE)
   vec1 <- rbeta(1E5, 2 + 10, 8 + 30)
   vec2 <- rbinom(1E5, size = 1, prob = vec1)
   prop.table(table(vec2))
   ```

   (c)
   ```
   vec1 <- runif(1E5, 0.35, 0.65)
   vec2 <- rgeom(1E5, prob = vec1)
   plot(table(factor(vec2, levels = 0:max(vec2))), type = 'h')
   mean(vec2 > 5)
   hist(vec1[vec2 == 3], freq = FALSE, nclass = 30)
   ```

   (d)
   ```
   vec1 <- rgamma(1E6, shape = 5, rate = 2)
   vec2 <- rpois(1E6, lambda = vec1)
   vec3 <- vec1[vec2 == 5]
   hist(vec3, freq = FALSE, nclass = 30)
   vec4 <- rpois(1E6, lambda = vec3)
   plot(table(factor(vec4, levels = 0:max(vec4))), type = "h")
   ```

   (e)
   ```
   vec1 <- abs(rnorm(1E5, mean = 0, sd = 2))
   vec2 <- rpois(1E6, lambda = vec1)
   hist(vec1[vec2 == 5], freq = FALSE, nclass = 30)
   ```

4. Explique o algorítmo de *Naïve Bayes* para classificação, utilizando o contexto da Questão 2. Quais são as característica, vantagens e desvantagens deste método?

5. Explique distribuição amostral, função de verossimilhança e distribuição *a posteriori* e como estão relacionadas a *paradigmas* de inferência estatística.

6. **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. Qual a probabilidade de que ela seja apoiadora do "Labour"? E dos "LibDem"?

7. No contexto do problema anterior:

   (a) Escreva os dados e quantidades de interesse na notação geral de variável observada Y e parâmetro θ.

   (b) Forneça a distribuição a priori do problema.

   (c) Forneça a verossimilhança do problema.

   (d) Encontre a distribuição à posteriori.

   (e) Esboçe um gráfico com distribuições priori e posteriori

(f) Forneça a distribuição preditiva.

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

$$f(y|\theta) = \theta \exp\{-\theta y\}$$

Obtenha a priori de Jeffreys, a expressão da posteriori, da marginal e da distribuição preditiva.

9. No contexto do problema anterior imagine que está se medindo o tempo para atender clientes em um banco. Forneça as expressões caso seja assumida como priori uma distribuição gama de média 0,25 e desvio padrão 1. Além disto considere que foi observada uma amostra aleatória de 20 tempos de atendimento e o tempo médio dos dados desta amostra foi de 4,8 minutos.

10. Forneça os comandos para desenhar em um gráfico as curvas da priori, verossimilhança e posteriori do item anterior.

11. Um experimento envolvendo uma série de reações químicas, que resulta em obter ou não o produto desejado, foi repetido consecutivamente (de forma independente), até que o terceiro resultado positivo fosse obtido. Para isto, foi necessário a realização de 12 experimentos no total. Deseja-se fazer inferência sobre a proporção de séries de reações que resultam no produto desejado. Por não haver informação prévia, optou-se por utilizar a priori de Jeffreys. Obtenha as expressões da priori e da posteriori.

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

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

Encontre a priori de Jeffreys e a correspondente distribuição posteriori para θ.

13. Encontre a priori de Jeffreys para θ no modelo geométrico:

$$f(y|\theta) = (1-\theta)^{y}\theta \; ; \;\; y = 0, 1, 2, \ldots$$

e obtenha as expressões da posteriori e preditiva para uma amostra $y_1, \ldots, y_n$.

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 imprória. Obtenha os gráficos da posteriori e da preditiva posteriori.

14. 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 deseja ajustar. Explique como as análises devem ser conduzidas neste caso, descrevendo pelo menos duas abordagens.

Implemente as abordagens mencionadas. (sugestão: implementar ao menos a aproximação gaussiana e MCMC).

15. Vai se modelar o número diário de acidentes em um trecho de rodovia com o modelo Poisson com priori "half-normal". Mostre a expressão da posteriori (a uma constante) e indique se é (ou não) compatível com alguma distribuição conhecida. 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 a sua média e moda.

Após obter a solução escrita implemente em um código. Faça inicialmente a aproximação de 5 pontos e depois obtenha a discretização com mais pontos. Escreva também uma função para posteriori que inclua o cálculo da constante normalizadora por integração numérica.

16. Considere um modelo Gama para a distribuição de uma amostra aleatória de uma variável, ou seja,

$$f(y|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\}.$$

Considere ainda prioris também Gama, independentes, para cada parâmetro.

(a) Obtenha a expressão da posteriori conjunta, a uma constante de proporcionalidade.

(b) Indique como pode ser obtida a distribuição posteriori marginal de cada parâmetro, isto é, $f(\alpha|y)$ e $f(\beta|y)$.

(c) Obtenha a distribuição posteriori condicional de cada parâmetro, isto é, $f(\alpha|\beta, y)$ e $f(\beta|\alpha, y)$.

(d) Explique como seriam poderiam ser obtidas amostras da posteriori neste caso.

Escolha valores dos hiperparâmetros do modelo e produza gráficos das prioris $f(\alpha)$, $f(\beta)$ e da conjunta $f(\alpha, \beta)$. Use o modelo especificado para gerar uma amostra de $n = 12$ valores. Obtenha gráficos dos itens a) a c). Obtenha amostras da posteriori conforme item d). Faça o histograma das amostras obtidas e sobreponha com a densidade (marginal) teórica.

17. Foi visto que distribuições da família exponencial, isto é, podem ser escritas na forma

$$f(y|\theta) = h(y)g(\theta)\exp\{t(y)c(\theta)\},$$

possuem prioris conjugadas. Use este resultado para obter, a partir da notação de família exponencial, a priori conjugada para uma amostragem de $Y \sim \text{Gama}(a, \theta)$ em que a é conhecido e θ é o parâmetro desconhecido sobre o qual deseja-se fazer inferência. Quais as quantidades amostrais (suficientes) necessárias para inferência?

18. Seja $Y|\theta \sim \text{Bin}(n, \theta)$ a priori de θ uma mistura das distribuições $\text{Beta}(1, 3)$ e $\text{Beta}(5, 2)$ com pesos $2/3$ e $1/3$, respectivamente. Uma amostra forneceu $y = 3, n = 9$. Esboce os gráficos de cada um de seus componentes da priori. Obtenha a expressão da posteriori e um esboço do seu gráfico.

Obtenha os gráficos para priori e suas componentes, da verossimilhança e da posteriori.

19. A classificação de qualidade de um componente elétrico é excelente $(y = 1)$, bom $(y = 2)$ ou ruim $(y = 3)$. A probabilidade de vários níveis de qualidade dependem da fábrica onde é produzido ($\theta_1 = 0$: fábrica A, $\theta_1 = 1$: fábrica B) e do tipo de máquina ($\theta_2 = 0$: máquina I, $\theta_2 = 1$: máquina II, $\theta_2 = 2$: máquina III). As probabilidades para $y = 2$ são dadas pela Tabela 3. Além disso, distribuição conjunta de (θ_1, θ_2) é dada na Tabela 4. Encontre a distribuição posteriori conjunta de $\theta_1, \theta_2|y = 2$ e cada uma das distribuições marginais. Tendo observado $y = 2$: (a) qual a combinação fábrica/máquina é a mais provável de ter produzido esse componente? (b) Qual a posteriori (conjunta)? (c) Quais são as marginais? (d) E as condicionais?

Tabela 1: Probabilidades condicionais $y = 2$

$\Pr(y = 2\|\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0,3	0,2
$\theta_2 = 1$	0,4	0,2
$\theta_2 = 2$	0,6	0,3

Tabela 2: Probabilidade a priori das combinações fábrica/máquina

$\Pr(\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0,1	0,2
$\theta_2 = 1$	0,2	0,3
$\theta_2 = 2$	0,1	0,1

20. Um conjunto $y_1, y_2, \ldots, y_n$ é `permutável` (*exchangeable*) se a probabilidade conjunta $f(y_1, y_2, \ldots, y_n)$ é invariante a uma permutação π dos índices, ou seja, $f(y_1, y_2, \ldots, y_n) = f(\pi(y_1), \pi(y_2), \ldots, \pi(y_n))$ para qualquer permutação π dos índices.

Um conjunto $y_1, y_2, \ldots, y_n$ é `independente` se a probabilidade conjunta $f(y_1, y_2, \ldots, y_n)$ é fatorável $f(y_1, y_2, \ldots, y_n) = f(y_1) \cdot f(y_2) \cdot f(y_n)$. Desta forma, para $n = 2$ isto implica que $f(y_2|y_1) = f(y_2)$

Considere uma urna com uma bola branca e uma preta com probabilidade $1/2$ de retirar cada uma delas. As bolas serão retiradas da urna sem reposição. Mostre que as variáveis definidas pelos resultados das retiradas são permutáveis, mas não independentes.

Dica de notação: Denote a primeira bola ser branca por $Y_1 = 0$ e preta por $Y_1 = 1$. Use análogo para segunda bola.

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

   (a) A qual curso voce arriscaria dizer que o estudante sorteado pertence? Justifique.

   (b) Resolva a 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

22. 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*) de probabilidades para os tempos de atendimento. Para conhecer melhor o processo vai se coletar uma amostra registrando o tempo gasto em n atendimentos.

   (a) Escreva a função de probabilidades para as observações.

   (b) Escreva a função de verossimilhança.

   (c) Reconheça uma distribuição de probabilidades proporcional à esta expressão da verossimilhança. Qual é a distribuição e com quais valores de parâmetros?

   (d) Forneça a expressão de uma distribuição à *priori* conjugada.

   (e) Qual(ais) valor(es) você especificaria para o(s) parâmetro(s) da distribuição *priori* conjugada (hiper-parâmetros)? Justifique.

   (f) Obtenha a expressão da distribuição marginal (preditiva a priori).

   (g) Explique como faria para simular valores desta distribuição marginal.

   (h) Obtenha a expressão da posteriori e indique qual a distribuição de probabilidades.

   (i) Qual(ais) quantidade(s) calculada(s) com a amostra são necessárias para a obtenção da verossimilhança e posteriori?

   (j) O estudo foi feito em uma primeira agência e o tempo médio observado para atender 20 clientes foi de 10,5 minutos. Forneça a expressão da distribuição a posteriori para θ com estes valores.

   (k) Qual o valor esperado da distribuição a posteriori para θ?

   (l) Mostre que o valor esperado da posterior é uma média ponderada entre o valor esperado da priori e do estimador de máxima verossimilhança (inverso da média dos dados).

   (m) Qual seria a distribuição preditiva para uma nova observação após obter esta primeira amostra?

   (n) Qual o valor esperado da distribuição marginal $[Y]$?

   (o) Em uma segunda agência, o estudo foi feito com 12 clientes e o tempo médio de atendimento foi de 8,5 minutos. Qual a posteriori após considerar os dados desta segunda agência?

   (p) Após algum tempo decidiu-se por igualar os tamanhos das amostras e completou-se a amostra da segunda agência anotando-se os tempos de atendimento de mais oito clientes que totalizaram 108 minutos. Compare os resultados da primeira agência e da segunda após esta complementação da amostra.

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

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

24. Comente sobre intervalos HPD frequentemente usados em inferência bayesiana em contraste com intervalos assintóticos e baseados em quantis, apontando as principais características e diferenças.

25. 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. Explique com detalhes como as análises devem ser conduzidas neste caso, considerando pelo menos duas abordagens (de natureza distinta).

26. Suponha que em um experimento físico, a variação temporal de uma corrente elétrica, denotada por X, tem distribuição uniforme no intervalo $[0, \theta]$. Sendo $X_1, \cdots, X_n$ realizações aleatórias desse experimento e assumindo o seguinte modelo a priori para θ

$$\pi(\theta) = \frac{10}{\theta^{11}} \underset{[1,\infty)}{\mathbb{I}(\theta)},$$

   pede-se

   a) Mostre que a distribuição a posteriori de θ é dada por

$$\pi(\theta|X_1, \cdots, X_n) = \frac{(n+10)\theta_0^{n+10}}{\theta^{n+11}} \underset{[\theta_0,\infty)}{\mathbb{I}(\theta)},$$

   sendo $\theta_0 = \max\left\{X_{(n)}, 1\right\}$ para $X_{(n)} = \max\{X_1, \cdots, X_n\}$.

   b) Obtenha o estimador de Bayes sob perda $0 - 1$ de θ.
   Dica: *Qual o valor de θ que fornece o maior valor de $\pi(\theta|X_1, \cdots, X_n)$?*

   c) Obtenha o intervalo HPD com 90% de credibilidade de θ.
   Dica: *Note que $\pi(\theta|X_1, \cdots, X_n)$ é unimodal e decrescente em θ.*

27. Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição com função de probabilidades:

$$f(x|\theta) = \theta \exp\{-\theta x\}, \quad \text{com } x > 0 \text{ e } \theta > 0.$$

   Suponha o modelo a priori de Jeffreys para θ (ou seja, $\pi(\theta) \propto 1/\theta$, $\theta > 0$). Com isso, considere o teste de hipóteses: $\mathrm{H}_0 : \theta \leq 1$ vs $\mathrm{H}_1 : \theta > 1$.

   (a) Obtenha o modelo a posteriori de θ, e expresse o estimador de Bayes sob perda quadrática para θ.

   (b) Supondo $\sum_{i=1}^{n} x_i = 1,5$, decida sobre H_0 contra H_1 considerando a seguinte estrutura de perda

$L(\theta, a)$	Decisão	
	$\theta \leq 1$	$\theta > 1$
$\theta \leq 1$	0	1
$\theta > 1$	1	0

   Considere $n = 10$ e expresse os comandos em R para obter a resposta.

28. Em algoritmos de MCMC, para se verificar se as cadeias atingiram a distribuição estacionária (convergência), é útil verificar se as diferentes cadeias já se misturaram a ponto de se tornarem indistinguíveis. Para isso é necessário dispor de múltiplas cadeias e utilizar pontos de partida bastante dispersos. O teste de Gelman-Rubin calcula a estatística $\hat{R}$ (`Rhat`), definida como

$$\hat{R} = \frac{S}{W}$$

sendo S o desvio padrão calculado a partir da mistura de todas as cadeias, e W a média dos desvios padrão das cadeias indivíduais. Valores de $\hat{R}$ muito maiores do que 1 são indicação de que ainda não houve convergência. A medida que as cadeias evoluem, o valor de $\hat{R}$, tende a se aproximar de 1. Como regra geral, considera-se aceitável valores $\hat{R} < 1.05$. Outra questão pertinente é saber se a informação contida na amostra simulada é suficiente. Ou seja, desejamos saber se o tamanho n_s da amostra simulada da distribuição posterior de fato tem a acurácia de uma amostra de n elementos independentes. Essa questão se justifica pelo fato de que os valores gerados nas cadeias tendem a apresentar autocorrelação. Uma forma de verificar esse fato se dá por meio do cálculo do número efetivo de amostras (`n.eff`)

$$n_{eff} = \frac{1}{1 + 2\sum_{i=1}^{\infty} \rho_k}$$

onde ρ_k são as correlações para o espaçamento de k passos na cadeia. Quanto mais próximo n_{eff} estiver de n, melhor.

O resultado abaixo é o resumo da distribuição posterior para o parâmetro θ de uma distribuição Binomial com $n = 3000$, obtido via JAGS

```
          mean    sd  2.5%   50% 97.5% [...]  Rhat n.eff
theta    0.163 0.136 0.006 0.129 0.508 [...] 1.12    752
```

Com esse resultado, comente sobre os valores obtidos de `Rhat` e `n.eff`, e o que deveria ser feito para que esses valores fiquem mais adequados.

29. Descreva o modelo sendo ajustado, distribuições envolvidas, parâmetros e constantes. Esboce ainda a estrutura dos dados nas seguintes declarações de modelos em JAGS.

(a)
```
model{
  for (i in 1:N){
    x[i] ~ dbern(p)
  }
  p ~ dbeta(alpha, beta)
  alpha <- 1
  beta <- 1
}
```

(b)
```
model{
  for(i in 1:M){
     for(j in 1:N){
        y[i,j] ~ dnorm(mu[i], tau)
     }
     mu[i] ~ dnorm(theta, tauD)
  }
  tau <- pow(sigma, -2)
  sigma ~ dunif(0, 100)
  theta ~ dnorm(0, .001)
  tauD <- pow(delta, -2)
  delta ~ dunif(0, 100)
}
```

(c)
```
model{
  for (i in 1:N){
    y[i] ~ dbern(p[i])
    logit(p[i]) <- a[g[i]] * x[i]
  }
  for (j in 1:K){
    a[j] ~ dnorm(mu.a, tau.a)
  }
  mu.a ~ dnorm(0, 0.0001)
  tau.a <- pow(sigma.a, -2)
```

```
    sigma.a ~ dunif(0, 1000)
  }
```

(d)
```
model{
    for (i in 1:N){
      y[i] ~ dnorm(mu[i], tau)
      mu[i] <- a[g[i]] * x[i] + b[g[i]]
    }
    for (j in 1:K){
      a[j] ~ dnorm(mu.a, tau.a)
      b[j] ~ dnorm(mu.b, tau.b)
    }
    mu.a ~ dnorm(0, 0.0001)
    mu.b ~ dnorm(0, 0.0001)
    tau <- pow(sigma, -2)
    sigma ~ dunif(0, 1000)
    tau.a <- pow(sigma.a, -2)
    tau.b <- pow(sigma.b, -2)
    sigma.a ~ dunif(0, 1000)
    sigma.b ~ dunif(0, 1000)
  }
```

(e)
```
model{
    for (t in 1:T) {
      y[t] ~ dbeta(a[t], b[t])
      a[t] <- mu[t] * phi
      b[t] <- (1 - mu[t]) * phi
      logit(mu[t]) <- alpha + beta * x[t]
    }
    alpha ~ dnorm(0, 10^-2)
    beta ~ dnorm(0, 10^-2)
    phi ~ dunif(0, 10)
  }
```

(f)
```
model
  {
    for (i in 1:T) {
      y[i] ~ dbin(p[i], K)
      logit(p[i]) <- alpha + beta_1 * x_1[i] + beta_2 * x_2[i]
    }
    alpha ~ dnorm(0.0,0.01)
    beta_1 ~ dnorm(0.0,0.01)
    beta_2 ~ dnorm(0.0,0.01)
  }
```

(g)
```
model{
    for(i in 1:n){
        y[i] ~ dpois(mu[i])
        mu[i] <- ifelse(i > k, lambda, theta)
    }
    k ~ dcat(ano[]) # distribuição para dados categóricos ou discretos
    theta ~ dgamma(a1, b1)
    lambda ~ dgamma(a2, b2)
    dif <- lambda - theta
    ib1 ~ dgamma(c1, d1)
    b1 <- 1/ib1
    ib2 ~ dgamma(c2, d2)
    b2 <- 1/ib2
```

```
    a1 <- 0.5
    a2 <- 0.5
    c1 <- 0.1
    c2 <- 0.1
    d1 <- 1
    d2 <- 1
  }
```

30. Com base no que você viu no curso até agora, responda:

    (a) Usando apenas palavras, como você explicaria para um leigo o que é inferência bayesiana?

    (b) Usando notação estatística, como você explicaria para um estatístico (mas que não sabe ainda) o que é inferência bayesiana?

    (c) 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 sob os parâmetros? Como os dados são tratados? Como as informações prévias podem ser consideradas?)

    (d) O que você julga serem os pontos fortes e fracos da inferência bayesiana?

31. Decidiu-se examinar um conjunto de imagens médicas tomadas independentemente de diferentes indivíduos. O objetivo é fazer inferências sobre a proporção de imagens com determinada característica morfológica. Em particular deseja-se avaliar se a proporção de indivíduos com alteração está abaixo de 40%.

    - Em um primeiro estudo tomaram-se 9 imagens sendo detectadas 3 com alteração.
    - Em um segundo estudo decidiu-se tomar-se imagens até que se obtivesse 3 com alteração. Ocorreu que foram tomadas 9 imagens no total.

    (a) Como seria testada a hipótese de interesse em cada estudo no enfoque não Bayesiano?

    (b) Neste caso os dois estudos forneceriam a mesma conclusão estatística?

    (c) Em uma análise Bayesiana com uma priori $[\theta] \sim \text{Be}(1,5; 1,5)$, qual seria a posteriori em cada estudo?

    (d) Como deveria ser avaliada a posteriori para concluir sobre a hipótese de interesese? Os estudos produziriam a mesma conclusão estatística? Justifique.

    (e) Se alternativamente decide-se adotar a priori de Jeffreys em cada estudo, elas seriam as mesmas a produziram a mesma posteriori? Justifique.

32. 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 a 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*.

33. 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.

    (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 dada pelos valores a seguir, e que $\mu = 10$. Obtenha a expressão da posteriori.

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

34. Vimos em aula o exemplo de taxas bayesianas baseadas na distribuição de Poisson. Vamos agora considerar um exemplo semelhante, porém com a distribuição Binomial. Considere que temos uma amostra de N grupos, cada um com n_i indivíduos onde conta-se o número y_i com determinada característica. Como contexto ilustrativo considere que registra-se o números de aprovados em diferentes turmas de uma disciplina.

   (a) Escreva o modelo, adotando a priori conjugada. Explique cada termo e identifique o parâmetro de interesse.

   (b) Obtenha a posteriori.

   (c) A distribuição posteriori é a inferência completa sobre o parâmetro. Entretanto se for necessário um relato resumindo a informação final (posteriori) em um único valor qual(quais) seriam as opções?

   (d) Caso desejar-se reportar a posteriori por um intervalo de valores, como seriam obtidos os limites de tal intervalo? Forneça ao menos duas maneiras explicando a forma de obtenção e comentando a diferença entre elas. Explique ainda por que este intervalo **não** é chamado de *intervalo de confiança*.

   (e) Explique como obter (segundo a análise bayesiana) a predição do número de aprovados para uma novo grupo com n_p indivíduos.

   (f) Explique como o procedimento bayesiano empírico seria efetivado neste caso.

35. Descreva o modelo sendo ajustado e a estrutura dos dados que seria necessária nas seguintes declarações de modelos em JAGS.

   (a)
```
model{
  for (i in 1:N){
    x[i] ~ dbern(p)
  }
  p ~ dbeta(alpha, beta)
  alpha <- 1
  beta <- 1
}
```

   (b)
```
model{
  for(i in 1:M){
     for(j in 1:N){
        y[i,j] ~ dnorm(mu[i], tau)
     }
     mu[i] ~ dnorm(theta, tauD)
  }
  tau <- pow(sigma, -2)
  sigma ~ dunif(0, 100)
  theta ~ dnorm(0, .001)
  tauD <- pow(delta, -2)
  delta ~ dunif(0, 100)
}
```

   (c)
```
model{
for(i in 1:n){
    Y[i] ~ dpois(mu[i])
    mu[i] <- N[i] * lambda
   }
   lambda ~ dgamma(a, beta)
   tau ~ dgamma(c, d)
   beta <- 1/tau
   a <- 1.2
   c <- 1.5
   d <- 2
}
```

   (d)
```
model{
for(i in 1:n){
    Y[i] ~ dbern(q[i])
```

```
    logit(q[i]) <- beta[1] + beta[2]*X[i,1] + beta[3]*X[i,2]
   }
   for(j in 1:3){
    beta[j] ~ dnorm(0,0.1)
   }
}
```

36. Considere o problema de uma regressão linear simples usual com o regressor x e a variável resposta Y. Deseja-se implementar um algorítmo para análise bayesiana deste modelo. Descreva os passos que seriam necessários e obtenha as expressões relevantes para implementação do algorítmo. Inclua no algoritmo a possibilidade de predição. Faça escolhas quando necessário como, por exemplo, de priori.

37. Considere o seguinte problema:

    Exemplo (aproveitamento de saques em jogos de tênis) *Considere dados (iid) $y = (y_1, \ldots, y_n)$ das taxas de sucesso no primeiro saque de um jogador de tênis em n jogos de um campeonato. Assuma o modelo $[Y|\theta] \sim f(y_i|\theta) = \theta(\theta+1)y_i^{\theta-1}(1-y_i)$ com $y_i \in (0,1)$ e $\theta > 0$. Não existe uma família conjugada usual para este modelo e adota-se uma priori gama $[\theta] \sim \mathrm{G}(1,1)$.* Foram obtidos dados tal que $n = 20$, $\sum_i \log(y_i) = -4,59$. Obtenha a expressão da posteriori e discuta como obter inferências (bayesianas) de interesse, incluindo resumos da posteriori e preditivas. Discuta e explique procedimentos e passos adotados em situações em que não há expressões analíticas (fechadas).

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

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

    e assume-se como distribuição a priori

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

    Obtenha a expressão da posteriori e da distribuição preditiva.

39. A distribuição de falhas ao longo do comprimento de uma fibra artificial segue um processo de Poisson, de modo que o número de falhas de um comprimento l da fibra é $\mathrm{Po}(l\theta)$. Pouco se sabe sobre θ. O número de falhas obtidos em 5 de fibras de comprimentos de 10, 15, 25, 30 e 40 metros, respectivamente, foram 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.

40. 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.

41. Considere o problema de uma regressão linear simples usual com o regressor x e a variável resposta Y. Deseja-se implementar um algorítmo para análise bayesiana deste modelo. Descreva os passos que seriam necessários e obtenha as expressões relevantes para implementação do algorítmo. Inclua no algoritmo a possibilidade de predição. Faça escolhas qdo necessário como, por exemplo, de priori.

42. Foi visto no curso e em provas anteriores 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 $\log(\lambda) \sim \mathrm{N}(a, b)$. É possível obter uma posteriori em forma fechada e conhecida neste caso? Qual a expressão da posteriori (se necessário a uma constante de proporcionalidade). Se voce tivesse dados de contagem para analisar sob o paradigma Bayesiano com este modelo, como procederia? Qual ou quais procedimentos adotaria para obter inferências à posteriori?

43. Encontre a priori de Jeffreys para θ no modelo geométrico:

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

para uma amostra $x_1, \ldots, x_n$. Obtenha as expressões da posteriori e da preditiva.
Considere uma amostra de dados: $5, 4, 7, 3, 2, 2, 5, 6, 8, 10$ Forneça a expressão da posteriori e o um valor predito para uma nova observação.

44. Considere o problema de uma regressão linear simples usual com o regressor x e a variável resposta Y. Deseja-se implementar um algorítmo para análise bayesiana deste modelo. Descreva os passos que seriam necessários e obtenha as expressões relevantes para implementação do algorítmo. Inclua no algoritmo a possibilidade de predição. Faça escolhas qdo necessário como, por exemplo, de priori.

45. A classificação de qualidade de um componente elétrico é excelente ($y = 1$), bom ($y = 2$) ou ruim ($y = 3$). A probabilidade de vários níveis de qualidade dependem da fábrica onde é produzido ($\theta_1 = 0$: fábrica A, $\theta_1 = 1$: fábrica B) e do tipo de máquina ($\theta_2 = 0$: máquina I, $\theta_2 = 1$: máquina II, $\theta_2 = 2$: máquina III). As probabilidades para $y = 2$ são dadas pela Tabela 3. Além disso, distribuição conjunta de (θ_1, θ_2) é dada na Tabela 4. Encontre a distribuição posteriori conjunta de $\theta_1, \theta_2|y = 2$ e cada uma das distribuições marginais. Tendo observado $y = 2$:

    (a) Qual a combinação fábrica/máquina é a mais provável de ter produzido esse componente?
    (b) Qual a distribuição posteriori (conjunta)?
    (c) Quais são as distribuições marginais?
    (d) E as condicionais?
    (e) Esboce gráficos para os itens anteriores

46. 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 $y_1, y_2, \ldots y_n$ 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 τ?
    (d) Qual é a distribuição posteriori para a priori de Jeffreys?

47. Seja uma variável aleatória Y e o seguinte modelo de probabilidades: $Y|\theta \sim \text{Bin}(n, \theta)$. Adota-se como priori de θ uma mistura das distribuições $\text{Beta}(2, 10)$ e $\text{Beta}(4, 2)$ com pesos 0,20 e 0,80, respectivamente. Uma amostra forneceu $y = 6, n = 20$. Esboce os gráficos de cada componentes da priori. Obtenha a expressão da posteriori e um esboço do seu gráfico.

Tabela 3: Probabilidades condicionais $y = 2$

$\Pr(y = 2\|\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0,3	0,2
$\theta_2 = 1$	0,4	0,2
$\theta_2 = 2$	0,6	0,3

Tabela 4: Probabilidade a priori das combinações fábrica/máquina

$\Pr(\theta_1, \theta_2)$	$\theta_1 = 0$	$\theta_1 = 1$
$\theta_2 = 0$	0,1	0,2
$\theta_2 = 1$	0,2	0,3
$\theta_2 = 2$	0,1	0,1

48. 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.

49. Seja $Y \sim \text{Gamma}(\alpha, \lambda)$ com priori (independente) $\alpha \sim \text{Gama}(a, b)$ e $\lambda \sim \text{Gamma}(c, d)$ em que λ, b e d são parâmetros de taxa.

(a) Obtenha a expressão da posteriori $f(\alpha, \lambda|y)$.

(b) Obtenha a expressão da posteriori $f(\lambda|\alpha, y)$.

(c) Obtenha a expressão da posteriori $f(\alpha|\lambda, y)$.

(d) Obtenha a expressão da posteriori $f(\lambda|y)$.

(e) Obtenha a expressão da posteriori $f(\alpha|y)$.

(f) Explique como obter simulações de cada uma das distribuições acima.

(g) Explique como obter simulações da marginal (preditiva a priori).

(h) Explique como obter simulações da marginal (preditiva a posteriori).

(i) Foi proposto o seguinte algorítmo para simular da posteriori conjunta:

  i. Escolha um valor inicial para α.

  ii. Simule um valor de $f(\lambda|\alpha, y)$, em que α é o último valor obtido.

  iii. Simule um valor de $f(\alpha|\lambda, y)$, em que λ é o último valor obtido.

  iv. Repita os dois passos anteriores N vezes.

  v. Os N pares α, λ são uma amostra de $f(\alpha, \lambda|y)$.

50. Um teste para detectar a presença ou não de certa doença é aplicado em uma população. Sabe-se que o teste acerta 90% dos que possuem a doença (*sensibilidade*) e 80% dos que não a possuem (*especificidade*). A ocorrência da doença na população alvo é de 2% (*prevalência*). O interesse é realizar uma análise bayesiana completa. Pede-se, explicando passo a passo:

  (a) **Análise da Posteriori:** Determinar a probabilidade de um indivíduo ter a doença, dado um resultado positivo (*valor preditivo positivo*), e a probabilidade de não ter a doença, dado um resultado negativo (*valor preditivo negativo*).

  (b) **Análise pelo Fator de Bayes:** Avaliar a força da evidência fornecida por um teste positivo em favor da hipótese de que o paciente tem a doença, em comparação com a hipótese de que não a tem.

  (c) **Análise pela Teoria da Decisão:** Suponha que a perda (ou custo) de um diagnóstico *falso negativo* (diagnosticar "não doente"quando a pessoa tem a doença) seja 5 vezes maior que a perda de um diagnóstico *falso positivo* (diagnosticar "doente"quando a pessoa não tem a doença). Qual seria a decisão ótima entre diagnosticar como 'Doente' ou 'Não Doente' para cada resultado do teste, a fim de minimizar a perda esperada?

  (d) Discuta os resultados.

  (e) Discuta as características de cada e as diferenças dos métodos.

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

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

53. Descreva o modelo sendo ajustado, distribuições envolvidas, parâmetros e constantes. Esboce ainda a estrutura dos dados nas seguintes declarações de modelos em JAGS.

  (a)
```
model{
  for (i in 1:N){
    x[i] ~ dbern(p)
  }
  p ~ dbeta(alpha, beta)
  alpha <- 1
  beta <- 1
}
```

(b)
```
model {
  for (i in 1:n) {
    y[i] ~ dnorm(0, tau)
  }
  tau ~ dgamma(a, b)
  sigma <- 1 / sqrt(tau)
}
```

(c)
```
model{
  for(i in 1:M){
     for(j in 1:N){
        y[i,j] ~ dnorm(mu[i], tau)
     }
     mu[i] ~ dnorm(theta, tauD)
  }
  tau <- pow(sigma, -2)
  sigma ~ dunif(0, 100)
  theta ~ dnorm(0, .001)
  tauD <- pow(delta, -2)
  delta ~ dunif(0, 100)
}
```

(d)
```
model {
    for (i in 1:n) {
        y[i] ~ dgamma(alpha, lambda)
    }
  alpha ~ dgamma(a, b)
  lambda ~ dgamma(c, d)
}
```

(e)
```
model {
  y ~ dbin(theta, n)
  k ~ dcat(p)
  theta ~ dbeta(alpha[k], beta[k])
}
```

(f)
```
model{
for(i in 1:n){
    y[i] ~ dpois(mu[i])
    mu[i] <- ifelse(i > k, lambda, theta)
}
k ~ dcat(ano[])
theta ~ dgamma(a1, b1)
lambda ~ dgamma(a2, b2)
dif <- lambda - theta
ib1 ~ dgamma(c1, d1)
b1 <- 1/ib1
ib2 ~ dgamma(c2, d2)
b2 <- 1/ib2
a1 <- 0.5
a2 <- 0.5
c1 <- 0.1
c2 <- 0.1
d1 <- 1
d2 <- 1
}
```