Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Bacharelado em Estatística
Inferência Bayesiana
Prof. Gustavo L. Gilardoni

Lista de Exercícios 2

Trata dos conceitos básicos associados ao paradigma Bayesiano: Cálculo da distribuição a posteriori usando a Fórmula de Bayes a partir da distribuição a priori e a verossimilhança; distribuições preditivas, estimação pontual e por intervalos e testes de hipóteses.

Os exercícios não especificam quando são requeridos cálculos exatos ou aproximados. Em cada caso deve-se decidir qual das duas possibilidades é mais conveniente.

Os exercícios numerados em verde são optativos.

1. O seu professor chega na sala de aula e mostra uma moeda. Você suspeita que a moeda possa ser falsa e ter duas caras. Considere a priori probabilidades iguais para os eventos da moeda ser falsa ou ser honesta (i.e. uma moeda bem equilibrada). **(i)** Calcule a sua probabilidade de obter cara num lançamento dessa moeda. **(ii)** Se o professor lançar a moeda e o resultado for cara, qual é agora a probabilidade dela ser falsa? **(iii)** Se o professor lançar a moeda n vezes e obter n caras, qual é a probabilidade dela ser falsa? Estude o comportamento desta probabilidade para n grande. **(iv)** Se o professor lançar a moeda uma vez e obter cara, qual é a probabilidade do próximo lançamento ser cara? **(v)** Explique porque é falso neste contexto a afirmação "os dois lançamentos da moeda são independentes", e explique qual seria a afirmação correta.

2. Seja $y_1, y_2, \ldots, y_n$ uma amostra da distribuição de Bernoulli com probabilidade de sucesso θ e considere uma distribuição a priori uniforme para θ. **(i)** Ache a distribuição a posteriori de θ e a sua média e variância. **(ii)** Mostre que é possível expressar a esperança a posteriori de θ da forma $(1-w)E(\theta)+w\hat{\theta}$, onde $E(\theta)$ e $\hat{\theta}$ são respectivamente a esperança a priori e a estimativa máximo verossímil de θ, e interprete este resultado. **(iii)** Se y_{n+1} é uma observação futura deste processo de Bernoulli, ache a distribuição preditiva $p(y_{n+1}|y_1, \ldots y_n)$.

3. Seja $y_1, y_2, \ldots, y_n$ uma amostra da distribuição de Bernoulli com probabilidade de sucesso θ e suponha que, a priori, $\eta = \text{logit}(\theta) = \log \frac{\theta}{1-\theta}$ segue uma distribuição Normal com média $\mu = 0$ e desvio padrão $\sigma = 10$. **(i)** Ache a densidade a priori de θ. **(ii)** No caso que $n = 12$ e $s = \sum_{i=1}^{12} y_i = 9$, calcule numericamente uma aproximação para a densidade a posteriori. **(iii)** No caso anterior, faça um gráfico comparando a distribuição a posteriori com a que seria obtida quando a priori $\theta \sim$ Uniforme(0,1) (i.é. a distribuição Beta com $\alpha = 10$ e $\beta = 4$).

4. No exercício 2, calcule **(i)** a estimativa bayesiana para Perda Quadrática e **(ii)** o limite da estimativa bayesiana para Perda Zero-Um quando $\epsilon \to 0$. No caso especial que $n = 12$, $s = \sum_{i=1}^{12} y_i = 9$, calcule **(iii)** a estimativa bayesiana sob Perda Absoluta e **(iv)** um intervalo HPD com nível 99%.

5. A função de perda $L(d, g(\theta)) = w(\theta)[d - g(\theta)]^2$, onde $w(\theta) \geq 0$ é chamada *Perda Quadrática Ponderada* (PQP, sem trocadilho). **(i)** Mostre que, quando existe, o estimador bayesiano para PQP é

$$\hat{g} = \frac{\int w(\theta)\, g(\theta) p(\theta \,|\, x)\, d\theta}{\int w(\theta)\, p(\theta \,|\, x)\, d\theta} = \frac{\mathrm{E}[w(\theta)\, g(\theta) \,|\, x]}{\mathrm{E}[w(\theta) \,|\, x]}\,.$$

(ii) Mostre que essa estimativa é a mesma que obteríamos usando PQ mas com a distribuição a priori $p_*(\theta) = w(\theta)\, p(\theta) / \int w(\theta)\, p(\theta)\, d\theta$ e interprete esse resultado.

6. Suponha que um experimento binomial fornece s sucessos em n ensaios, e que a priori a probabilidade de sucesso $\theta \sim$ Beta(α,β). **(a)** Calcule a estimativa bayesiana de θ sob a PQP $L(\theta, \hat{\theta}) = \theta^a\, (1-\theta)^b\, (\hat{\theta} - \theta)^2$; **(b)** Considere detalhadamente e interprete os casos (i) $a = b = -1$, (ii) $a = -1$, $b = 0$ e (iii) $a = 0$, $b = -1$;

7. (Gelman, Carlin, Stern e Rubin). Suponha que $x_1, \ldots, x_5 \overset{iid}{\sim}$ Cauchy(θ,1) (i.é. $p(y_1|\theta) = \pi^{-1}/[1 + (y_1 - \theta)^2]$ para $-\infty < y_1 < \infty$). Assuma que a sua distribuição a priori para θ é Uniforme(0,1). Dadas as observações $(x_1, \ldots, x_5) = (-2, -1, 0, 1.5, 2.5)$: **(i)** Compute a densidade a posteriori não normalizada [isto é, $p(\theta)\, p(x_1, \ldots, x_5|\theta)$] numa grade de pontos $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1$ para algum valor grande de m. **(ii)** Usando essa grade, calcule uma aproximação para $p(\theta|x_1, \ldots, x_5)$ e faça um gráfico dela. **(iii)** Amostre 1000 observações da distribuição a posteriori aproximada e faça um histograma dessa amostra. **(iv)** Usando a amostra da parte **(iii)**, amostre 1000 observações da distribuição preditiva de uma observação futura y_6. Calcule a média e faça um histograma dessa amostra.

8. Suponha que (x_1, x_2, x_3) dado p_1, p_2, p_3 segue uma distribuição Multinomial com parâmetros n e (p_1, p_2, p_3), onde $p_i \geq 0$ e $p_1 + p_2 + p_3 = 1$, e que, a priori, (p_1, p_2, p_3) segue uma distribuição de Dirichlet com parâmetros $(\alpha_1, \alpha_2, \alpha_3)$. **(i)** Ache a distribuição a posteriori de p_1, p_2, p_3 e as distribuições a posteriori marginais de p_i $(i = 1, 2, 3)$ e **(ii)** Calcule as estimativas bayesianas de p_i e de $p_j - p_i$ sob Perda Quadrática $(i, j = 1, 2, 3, i \neq j)$.

9. Na véspera do primeiro turno para a eleição de governador do DF de 2010, a Datafolha divulgou uma pesquisa indicando que, de 891 eleitores entrevistados que já tinham decidido em quem votar, Agnelo Queiroz tinha a preferência de 467, Weslian Roriz a de 315 e outros candidatos a de 109 eleitores. Formule um modelo para analizar esses dados. Ache estimativas bayesianas e construa intervalos críveis para **(a)** a proporção de votantes de Agnelo Queiroz e **(b)** a diferença entre a proporção de votantes de Agnelo e de Weslian Roriz.

10. Os dados a seguir mostram o resultado de 2 ensaios clínicos realizados nos anos 80 para estudar se o consumo diário de aspirina tem algum efeito na redução da taxa de mortalidade devido a doenças cardiacas. Assuma independência condicional tanto entre ensaios quanto entre os grupos placebo/aspírina dentro do ensaio e considere o modelo $x_{A,i}|\theta_{A,i} \sim$ Binomial $(n_{A,i};\theta_{A,i})$; $x_{P,i}|\theta_{P,i} \sim$ Binomial $(n_{P,i};\theta_{P,i})$.

	Aspirina		Placebo	
	Pacientes $(n_{A,i})$	Mortes $(x_{A,i})$	Pacientes $(n_{P,i})$	Mortes $(x_{P,i})$
Ensaio 1	810	85	406	52
Ensaio 2	832	102	850	126

(a) Construa uma dsitribuição a priori para o vetor de parâmetros $(\theta_{A,1}, \theta_{P,1}, \theta_{A,2}, \theta_{P,2})$. Explique. **(b)** Usando a priori da parte (a), ache uma estimativa bayesiana e construa um intervalo crível para a diferença entre os efeitos netos da aspirina nos dois ensaios (o efeito neto da aspirina no ensaio i é definida como $\eta_i = (\theta_{A,i} - \theta_{P,i}))$. **(c)** Usando a priori da parte (a), ache uma estimativa bayesiana e um intervalo crível para a diferença entre os efeitos brutos da aspirina nos dois ensaios (o efeito bruto da aspirina no ensaio i é $\theta_{A,i}$).

11. É conhecido que 25% dos pacientes de um certo grupo que sofrem de enxaqueca melhoram após duas horas de serem tratados com um *placebo*. Para verificar se uma droga nova é melhor que o placebo, $n = 20$ pacientes foram tratados com o placebo e verificou-se que após duas horas $s = 8$ deles relataram ter melhorado. Seja θ a probabilidade de um paciente tratado com a droga nova melhorar após duas horas. **(i)** Especifique a hipótese nula H_0 e a alternativa H_1; **(ii)** Usando a distribuição a priori "não informativa" $\theta \sim$ Uniforme(0,1), calcule as chances relativas a priori e a posteriori de H_1 e o correspondente Fator de Bayes; **(iii)** Seja $d = 1$ a decisão de rejeitar H_0 e $d = 0$ a de não rejeitar. Considere a função de perda de Neyman para a qual é 5 vezes mais custoso rejeitar H_0 quando ela é verdadeira do que não rejeitar quando ela é falsa [isto é, $L(d = 1, \theta \in H_0) = 5\, L(d = 0, \theta \notin H_0)$, $L(d = 1, \theta \notin H_0) = L(d = 0, \theta \in H_0) = 0$]. Calcule a decisão ótima a posteriori; **(iv)** É razoável chamar essa distribuição a priori de "não informativa" nesse problema? Se a sua resposta for negativa, sugira uma outra distribuição a priori e refaça os cálculos anteriores.

12. Considere o teste de uma hipótese simples $H_0 : \theta = \theta_0$ contra a alternativa $H_1 : \theta = \theta_1$ (i.é., o espaço paramétrico é $\Theta = \{\theta_0, \theta_1\}$!) O Lemma de Neyman-Pearson (N-P) afirma que as regiões críticas *admissíveis* são da forma

- $p(x \mid \theta_1) > k\, p(x \mid \theta_0)$ implica que $x \in C$ e
- $p(x \mid \theta_1) < k\, p(x \mid \theta_0)$ implica que $x \notin C$;

no sentido que essas regiões críticas minimizam a probabilidade do *Erro Tipo* II (β) para um valor fixo do *Erro Tipo* I (α). Mais precisamente, para qualquer outra região crítica C_* tal que $\alpha_* = \mathbb{P}(x \in C_* \mid \theta_0) \leq \mathbb{P}(x \in C \mid \theta_0) = \alpha$, devemos ter que $\beta_* = \mathbb{P}(x \notin C_* \mid \theta_1) \geq \mathbb{P}(x \notin C \mid \theta_1) = \beta$. É usual nas aplicações escolher (fixar) o nível de significância α, embora na verdade é possível verificar que existe uma relação biunívoca entre α, β e o k usado na definição de C, no sentido que conhecendo o valor de qualquer um deles é possível determinar o valor dos outros dois.

Considere $X_1, \ldots, X_n \mid \theta \sim$ Exponencial (θ) (i.é., $f(x \mid \theta) = \theta\, e^{-\theta\, x}$ para $x > 0$) e o problema de testar $H_0 : \theta = 1$ contra a alternativa $H_1 : \theta = 2$. **(i)** Para $n = 10$, ache o valor de β e o de k quando $\alpha = 0.9$, 0.95 e 0.99; **(ii)** Repita o item anterior para $n = 30$ e $n = 100$; **(iii)** Repita os cálculos anteriores mas agora fixando $k = 2$ (i.é., ache nesse caso os valores correspondentes de α e de β); **(iv)** Qual sería a interpretação de fixar $k = 2$ do ponto de vista Bayesiano?