Exercícios de Estatística Bayesiana
Ano lectivo 2012 / 2013

1. Sabe-se que cerca de 30% dos gémeos humanos são idênticos. Gémeos idênticos têm necessariamente o mesmo sexo - metade são rapazes e a outra metade são raparigas. Um quarto dos gémeos que não são idênticos são ambos do sexo masculino, 25 % são ambos do sexo feminino e os restantes 50% são mistos - uma rapariga e um rapaz. Imagine que está no hospital e o médico diz-lhe que é mãe (pai) de duas gémeas. Qual é a probabilidade de que as gémeas sejam idênticas?

2. Muitas mulheres fazem o teste de gravidez assim que há uma falha no período mestrual. Nessa altura precoce, a probabilidade de o diagnóstico ser correcto se a mulher estiver efectivamente grávida, $P(+ \mid G)$, é de 0.983, e no caso de não estar grávida, $P(- \mid \bar{G})$, é 0.988. Sabe-se que a falha do período mestrual apenas em 12% dos casos corresponde de facto a gravidez, em 1.5% dos casos deve-se a alterações de saúde que requerem vigilância médica, e em 86.5% dos casos são meros atrasos sem qualquer significado.

   (a) Verifique que a probabilidade de resposta positiva no teste precoce de gravidez é aproximadamente igual a 0.13.

   (b) Calcule o valor de $P(G \mid +)$.

   (c) Se o resultado for negativo, nessa fase, qual é a probabilidade de afinal essa senhora estar grávida?

3. Um director comercial ao considerar a proporção de indivíduos de uma dada população que dariam preferência a um certo produto da empresa estabeleceu, para facilitar a análise, cinco casos aos quais atribuíu as seguintes probabilidades *a priori*:

Proporção de consumidores	Probabilidades *a priori*
0.10	0.25
0.15	0.35
0.20	0.20
0.25	0.15
0.30	0.05

   Através de um pequeno estudo de mercado, no qual foram inquiridos $n = 10$ indivíduos, verificou-se que havia $x = 2$ consumidores do produto em questão. Actualize as probabilidade dos cinco casos considerados na tabela.

4. Num certa pequena povoação, 1/3 dos dias do ano são chuvosos. O jornal local pretende prever se choverá ou não no dia seguinte. Sabe-se que três quartos dos dias chuvosos e 3/5 dos dias secos são correctamente previstos. Suponha que o jornal escreve que choverá no dia seguinte. Qual é a probabilidade que tal aconteça?

5. Na área florestal de um país do norte da europa ainda é possível encontrar linces. Em determinado momento do tempo, o número de linces existentes no local, X, pode variar entre 0 e 5, sendo

$$P(X = x) = \binom{5}{x} 0.6^x 0.4^{5-x},$$

para $x = 0, 1, \ldots, 5$. Planeia-se um estudo de forma a estudar o comportamento do lince no seu habitat. Infelizmente, verifica-se que é muito difícil visualizar um lince. Suponha, então, que o número de linces observado é descrito por uma v.a. Y com função massa de probabilidade (f.m.p.) dada por

$$P(Y = y \mid X = x) = \begin{cases} \binom{x}{y} 0.3^y 0.7^{x-y} & , 0 \leq y \leq x \\ 0 & , x < y \end{cases}$$

Obtenha a distribuição condicional de X, supondo que conseguiu avistar dois linces.

6. Uma certa espécie de peixes migra anualmente. Em determinado dia do ano, a probabilidade de que a migração se inicie é de 0.4. Nesse caso, um observador deverá espera T minutos antes de visualizar um peixe da espécie de interesse. Suponha que T é uma v.a. com distribuição exponencial de valor médio 20. Se a migração ainda não se tiver iniciado, então nenhum peixe será avistado. Obtenha a probabilidade de que a migração ainda não se tenha iniciado sabendo que não se avistou nenhum peixe após uma hora de observação.

7. Suponha que se $\theta = 1$ então a v.a. X tem distribuição $\mathrm{N}(1, \sigma^2)$ e se $\theta = 2$, X segue uma $\mathrm{N}(2, \sigma^2)$. Admita que $p(\theta = 1) = p(\theta = 2) = 0.5$. Sendo $\sigma = 2$, indique como calcular a f.d.p. marginal de X e represente-a graficamente.

8. Seja X a v.a. que conta o número de vezes que se obtém a face 6 ao lançar um dado 1000 vezes. Seja θ a probabilidade associada à saída da face 6. Suponha que se pensa que o dado não é equilibrado e que a probabilidade *a priori* de sair a face 6 é dada por:

$$p(\theta = 1/12) = 0.25, p(\theta = 1/6) = 0.50, p(\theta = 1/4) = 0.25.$$

Aproximando $f(x \mid \theta)$ pela distribuição normal, indique a expressão (aproximada) da distribuição marginal dos dados, $f(x)$ e represente-a graficamente.

9. Considere a variável aleatória Y que conta o número de caras que se obtêm ao lançar uma moeda ao ar n vezes. Considere que θ é a probabilidade de se observar uma "Cara".

   (a) Se a distribuição *a priori* para θ for uma distributição Uniforme no intervalo $(0,1)$, obtenha a distribuição dos dados:

   $$P(Y=y)=\int_0^1 P(Y=y\mid\theta)d\theta$$

   (b) Considere agora que a distribuição *a priori* para θ é uma distribuição Beta(α,β), $\alpha>0$ e $\beta>0$ com a seguinte função densidade de probabilidade:

   $$p(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\theta\in(0,1)$$

   Prove que a média *a posteriori* varia entre o valor médio da distribuição *a priori* e a frequência relativa do número de caras y/n. Para tal, determine o valor de λ tal que:

   $$E(\theta\mid y)=\lambda E(\theta)+(1-\lambda)\frac{y}{n}$$

   e verifique que $0<\lambda<1$.

   (c) No caso de que a distribuição *a priori* de θ seja uniforme, constate que a variância *a posteriori* de θ é sempre inferior à variância *a priori*.

10. Considere que a distribuição Beta(4,4) é apropriada para representar a probabilidade de que saia a face "cara"ao lançar ao ar uma moeda equilibrada. A moeda é lançada ao ar 10 vezes tendo-se observado que a face "cara" saiu menos de 3 vezes. Calcule a distribuição *a posteriori* de θ e represente-a graficamente, juntamente com a distribuiçã *a priori*.

11. A proporção (θ) da população adulta do estado da California (USA) que está de acordo com a pena de morte tem distribuição Beta(a,b), $a>0$ e $b>0$, com valor médio de 0.6 e desvio padrão igual a 0.3.

    (a) Determine os parâmetros da distribuição *a priori*. Represente graficamente a distribuição que obteve.

    (b) Suponha que numa amostra aleatória de 1000 californianos revelou uma aderência à pena de morte de aproximadamente 65%. Qual é a média e a variância da distribuição *a posteriori*?

12. Um médico decide testar um novo medicamento em 70 indivíduos que sofrem de uma certa doença. O tratamento é administrado durante seis meses. No fim desse período de tempo, verifica-se que $x = 34$ dos $n = 70$ indivíduos ainda sobrevivem. Seja θ a probabilidade de sobrevivência à doença.

    (a) Admita que a experiência clínica do médico lhe permite afirmar que a taxa de sucesso do medicamento deverá ser, em média, de 40%, com uma variância de 0.02. Indique a distribuição *a priori* para θ que resulta da informação fornecida pelo médico (use uma Beta(a,b), $a > 0$ e $b > 0$). Qual é a probabilidade *a priori* de que 50% dos indivíduos sobrevivam à doença?

    (b) Obtenha a distribuição *a posteriori* e calcule a média, a variância e a moda.

    (c) Qual é a probabilidade *a posteriori* de que 50% dos indivíduos sobrevivam à doença? Represente no mesmo gráfico as distribuições *a priori* e *a posteriori*. Poderá dizer que os dados são informativos relativamente a θ?

    (d) Assuma agora que se trata de uma doença pouco frequente e que o médico não consegue exprimir a sua opinião acerca de θ, antes de observar os dados. Qual será então a distribuição *a priori* para θ que deverá considerar? Que modificações se observam na distribuição *a posteriori*?

13. Considere que o número de defeitos existentes num rolo de fita magnética segue uma distribuição de Poisson(θ), $\theta > 0$. A distribuição *a priori* para θ é uma Gama(3, 1). Em cinco rolos de fita magnética observaram-se os seguintes números de defeitos: 2, 2, 6, 0 e 3, respectivamente para os rolos 1, 2, ...,5. Determine a distribuição *a posteriori* de θ.

14. Por volta dos anos 90, o *General Social Survey* reuniu informação acerca do nível de escolaridade, e correspondente número de filhos, de mulheres com aproximadamente 40 anos de idade (no momento de participação no estudo). Essas mulheres tinham à volta de 20 anos na década de 70, que foi um período de fertilidade baixa no EUA. Pretende-se avaliar o efeito da escolaridade no número de filhos. Sejam $(Y_{1,1}, Y_{2,1}, \ldots, Y_{n_1,1})$ e $(Y_{1,2}, Y_{2,2}, \ldots, Y_{n_2,2})$ as amostras correspondentes ao número de filhos das n_1 mulheres que têm bacharelato ou formação inferior e das n_2 que têm, no mínimo, bacharelato, respectivamente. Suponha que:

$$Y_{1,1}, Y_{2,1}, \ldots, Y_{n_1,1} \mid \theta_1 \sim i.i.d.\ Poisson(\theta_1), \theta_1 > 0$$

e

$$Y_{1,2}, Y_{2,2}, \ldots, Y_{n_2,2} \mid \theta_2 \sim i.i.d.\ Poisson(\theta_2), \theta_2 > 0.$$

Os dados observados foram:

$$n_1 = 111, \sum_{i=1}^{n_1} y_{i,1} = 217, \bar{y}_1 = 1.95$$

e

$$n_2 = 44, \sum_{i=1}^{n_2} y_{i,2} = 66, \bar{y}_2 = 1.50.$$

(a) Supondo que as distribuições *a priori* de θ_1 e θ_2 são $Gama(2, 1)$, obtenha as distribuições *a posteriori* de θ_1 e de θ_2.

(b) Represente graficamente as distribuições *a posteriori* de θ_1 e de θ_2 obtidas na alínea anterior. Indique os valores da média e da moda *a posteriori*.

(c) Obtenha os intervalos de credibilidade a 95%, com iguais probabilidades de cauda, para θ_1 e θ_2 *a posteriori*.

15. Considere que o tempo de vida das lâmpadas de um certo fabricante é adequadamente modelado por uma exponencial(θ), sendo θ desconhecido, cuja f.d.p. é:

$$f(x \mid \theta) = \theta e^{-\theta x}, x > 0.$$

Suponha que distribuição *a priori* para θ é uma $Gama(\alpha, \beta)$, para α e β conhecidos, e que o coeficiente de variação (CV) de θ toma o valor 0.5. Recolhe-se uma a.a. de dimensão n da produção diária da fábrica. Quantas lâmpadas se deverão escolher de forma a que a distribuição *a posteriori* apresente um CV de 0.1?

16. [1] Seja X uma v.a. que representa o tempo (em minutos) de atendimento no guichet de um consultório médico que é especificamente destinado a entregar resultados de exames. Suponha que X tem uma distribuição exponencial com função densidade de probabilidade dada por $f(x \mid \theta) = \theta e^{-\theta x}, x > 0$, sendo $\theta > 0$. Os tempos $x_1, x_2, ..., x_{20}$ necessários para servir 20 indivíduos foram registados, tendo-se obtido $\bar{x} = 3.8$ minutos. Considere que distribuição *a priori* de θ é uma gama de valor médio 0.2 e desvio padrão 1. Obtenha a distribuição *a posteriori* de θ. Indique os valores de $E(\theta \mid \mathbf{x})$ e de $Var(\theta \mid \mathbf{x})$. Represente graficamente as distribuições *a priori* e *a posteriori*.

17. À média (θ) de uma distribuição Normal de variância 72 é atribuída uma distribuição *a priori* Normal de média 800 e variância 12. Que dimensão deverá ter uma amostra da população de forma a garantir que a variância da distribuição *a posteriori* não seja superior a 1?

18. Ao descrever os resultados de uma investigação, um estatístico declarou que a distribuição *a posteriori* da média θ de uma distribuição normal com variância de 100 era ainda normal com média igual a 52 e variância de 10. Acrescentou ainda que a informação experimental consistiu numa amostra de 4 elementos com média 55. Especifique completamente a distribuição *a priori* de θ que se sabe também ser normal.

19. Suponha que se está a testar um novo medicamento para baixar a tensão arterial de doentes hipertensos. Para indivíduos que apresentam tensão arterial diastólica de 95mmHg, espera-se que o novo medicamento produza, em média, um decréscimo (θ) de 6mmHg e que a probabilidade de que θ exceda 12mmHg é de 15%. Assumindo uma distribuição *a priori* $\mathrm{N}(a, b^2)$ para θ, indique os valores de a e de b.

[1] Adaptado do exame 17 de Julho de 2006

20. Uma amostra aleatória de $n = 50$ adultos do sexo masculino é aleatoriamente seleccionada, tendo-se verificado que o peso médio dos indivíduos que constam da amostra é de 68 kg. Assuma que o peso X, em kg, é uma variável aleatória com distribuição Normal, de valor médio θ desconhecido e com desvio padrão de 10 kg, ou seja, $X \sim \mathrm{N}(\theta, 10^2)$. Suponha que a distribuição *a priori* para θ é uma N(80,18^2),

    (a) Indique a distribuição a posteriori de θ

    (b) Qual é a distribuição preditiva de um novo indíviduo da mesma população cujo peso é y^*?

21. [2] Vinte e um bébes foram seguidos desde a nascença de forma a registar a idade à qual proferiram a primeira palavra. Os tempos, em meses, que foram registados são os seguintes:

15	26	10	9	15	20
18	11	8	20	7	9
10	11	11	10	12	42
17	11	10			

Assuma que os dados provêm de uma distribuição $N(\theta, \sigma^2)$, $\theta \in \mathbb{R}$ e $\sigma > 0$ conhecido, tomando o valor $\sqrt{63}$, e que a distribuição *a priori* para θ é $N(b, d^2)$, $b \in \mathbb{R}$ e $d > 0$.

    (a) Elicite a distribuição *a priori* para a idade média (θ) de início da fala, sabendo que 25% das crianças falam pela primeria vez antes dos 8 meses e que apenas 9% excedem os 20 meses.

    (b) Indique a densidade *a posterior* de θ.

    (c) Represente graficamente as distribuições *a priori* e *a posterior*.

    (d) Calcule a probabilidade *a posteriori* de que θ seja superior a 2 anos.

    (e) Qual seria o valor da probabilidade da alínea anterior se tivesse usado uma distribuição *a priori* vaga para θ?

    (f) Nas condições da alínea (a), e considerando que vai observar mais $m = 10$ crianças, qual é a distribuição preditiva para a idade média na qual as crianças vão falar pela primeira vez?

22. Considere que $X_1, X_2, ..., X_n$ é um conjunto de n variáveis aleatórias independentes com distribuição Uniforme no intervalo $[0, \theta]$. Logo

$$f(x_i \mid \theta) = \frac{1}{\theta}, 0 < x_i < \theta.$$

Assuma que a distribuição *a priori* para θ é a seguinte distribuição imprópria:

$$p(\theta) = \frac{1}{\theta}, \theta > 0.$$

[2] Exame 27 Junho de 2006

(a) Prove que a distribuição *a posteriori* tem a seguinte expressão:

$$p(\theta \mid \mathbf{x}) = \frac{nM^n}{\theta^{n+1}}, \theta \geq M,$$

sendo $M = \max\{x_1, x_2, ..., x_2\}$.

(b) Prove que $E(\theta \mid \mathbf{x}) = \frac{n}{n-1}M$.

(c) Calcule o valor de $P\left(\theta > tM \mid \mathbf{x}\right)$ para $t \geq 1$.

23. Obtenha a distribuição *a priori* de Jeffreys das seguintes distribuições:

(a) Poisson(θ), $\theta > 0$.

(b) N(θ, σ^2), $\theta \in \mathbb{R}$ e $\sigma > 0$ conhecido.

(c) N(θ, σ^2), $\theta \in \mathbb{R}$ conhecido e $\sigma > 0$.

(d) N(θ, σ^2), $\theta \in \mathbb{R}$ e $\sigma > 0$, ambos desconhecidos.

(e) Geométrica(θ), cuja f.m.p. dada por $f(x \mid \theta) = (1-\theta)^{x-1}\theta, x = 1, 2, \ldots$ (Relembre que $E(X) = 1/\theta$).

24. [3] Considere que possui uma amostra aleatória de dimensão n, $\mathbf{x} = (x_1, x_2, ..., x_n)$, de uma distribuição com função densidade de probabilidade (f.d.p.) dada por

$$f(x \mid \theta) = \theta 3^{\theta} x^{-(\theta+1)}, x > 3, \theta > 0.$$

(a) Obtenha a distribuição *a priori* de Jeffreys para θ.

(b) Usando a distribuição obtida na alínea anterior, deduza a distribuição *a posteriori* de θ.

(c) Obtenha a média e a variância *a posteriori* de θ.

(d) Mostre que a distribuição preditiva de uma observação futura y do mesmo modelo f é dada por:

$$f(y \mid \mathbf{x}) = \frac{n\beta^n}{y(\beta - \ln 3 + \ln y)^{n+1}},$$

sendo $\beta = \sum_{i=1}^{n} \ln x_i - n \ln 3$.

25. Considere o modelo Binomial(n, θ). Escreva-o na forma da família exponencial e identifique o seu parâmetro natural (ϕ). Em seguida, prove que se $p(\phi) \propto 1$ então a distribuição *a priori* para θ é $p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$.

26. [4] Suponha que pretende estimar o parâmetro θ de uma distribuição *Binomial*(n, θ), sendo $0 < \theta < 1$. Considerando a distribuição *Beta*$(1, 1)$ como distribuição *a priori* para θ e a função perda

$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(1-\theta)},$$

mostre que a Regra de Decisão de Bayes é $a = d(x) = x/n$ para $0 < x < n$.

[3] Exame de recorrência - 2005/2006

[4] Exame de 1ª época - 2005/2006

27. [5] Uma criança em idade escolar faz um teste de forma a avaliar o seu coeficiente de inteligência. O resultado do teste é uma v.a. X com distribuição $N(\theta, 10^2)$, onde θ representa o verdadeiro coeficiente de inteligência da criança que é medido pelo teste. Suponha que, na população de crianças nas mesmas condições, a distribuição *a priori* para θ é $N(100, 15^2)$. Pretende-se decidir, com base nos resultados do teste, se se deve colocar a criança num grupo de leitura inicial, médio ou avançado. As acções possíveis são:

    a_1: colocar a criança num grupo inicial, i.e., decidir que $\theta \in \Theta_1 = (0, 90)$

    a_2: colocar a criança num grupo médio, i.e., decidir que $\theta \in \Theta_2 = (90, 110)$

    a_3: colocar a criança num grupo avançado, i.e., decidir que $\theta \in \Theta_3 = (110, \infty)$.

    Pensa-se que a seguinte função perda $L(\theta, a)$ poderá ser adequada:

$a \setminus \theta$	$\theta \in \Theta_1$	$\theta \in \Theta_2$	$\theta \in \Theta_3$
a_1	0	$\theta - 90$	$2(\theta - 90)$
a_2	$90 - \theta$	0	$\theta - 110$
a_3	$2(110 - \theta)$	$110 - \theta$	0

    Supondo que se observa $x = 115$, indique a melhor acção a tomar.

28. Considere que $X_1, X_2, ..., X_n$ é uma amostra aleatória de uma população com distribuição BinomialNegativa(r, θ), $r > 0$ e $0 < \theta < 1$, cuja função de probabilidade é dada por:

$$p(x \mid r, \theta) = \binom{x + r - 1}{x} \theta^r (1 - \theta)^x, x = 0, 1, 2, ...$$

    Assuma que r é conhecido.

    (a) Usando a distribuição conjugada natural de θ, obtenha a distribuição *a posteriori* de θ.

    (b) Suponha agora que recolheu uma amostra aleatória dessa população e que obteve $\sum_{i=1}^{n} x_i = 70$, sendo $r = 5$ e $n = 10$. Considere as hipóteses :
    $$H_0 : \theta \leq 0.5 \text{ vs. } H_1 : \theta > 0.5$$
    Qual das hipóteses apresenta uma maior probabilidade *a posteriori*, admitindo que possui uma informação "moderada"acerca de θ, a qual é introduzida no modelo por meio de uma distribuição Beta (2.0, 2.0)?

    (c) Indique o valor do factor de Bayes em favor de H_1. Poderá concluir que existe forte evidência em favor desta hipótese? (Recorde que se A e B forem inteiros então $F_{Beta(A,B)}(\theta_0) = 1 - F_{Binomial(A+B-1;\theta_0)}(A - 1)$).

[5] Trabalho 2005/2006 - exercício de treino

29. Dezasseis consumidores habituais de comida de uma cadeia de *fast food* foram recrutados para participar num estudo cujo objectivo era comparar o sabor de dois tipos de recheio usado num certo tipo de bolo confeccionado com carne de vaca. Um dos conjuntos de dezasseis recheios que foi analisado tinha sido congelado 8 meses atrás e mantido num congelador de elevada qualidade (com alterações de temperatura de menos do que um grau na escala de Fahrenheit) durante todo o período de tempo. Os restantes 16 recheios foram armazenados em congeladores de qualidade média, com variações de temperatura entre 0 e 15 graus Fahrenheit. Os gestores da cadeia de *fast food* pretendem verificar se a qualidade da congelação altera o sabor dos recheios e portanto avaliar se valerá a pena o elevado custo associado a uma congelação mais cuidada.
Os produtos são descongelados e cozinhados por um único **chef**. O planeamento da experiência é feito de forma a que seja "duplamente-cego", i.e., nem os empregados que servem o recheio nem os consumidores têm qualquer conhecimento sobre os produtos que estão a servir (consumir) em qualquer momento. No fim da experiência observa-se que 13 dos 16 consumidores preferiram os recheios que tinham sido congelados com maior cuidado (no congelador de elevada qualidade).

   (a) Qual é o modelo que sugere para analisar estes dados? Que pressupostos deve impor?

   (b) Seja θ a probabilidade de que os consumidores prefiram o produto mais caro. Sejam $Beta(0.5, 0.5)$, $Beta(1.0, 1.0)$ e $Beta(2.0, 2.0)$ três distribuições que reflectem diferentes níveis de conhecimento, *a priori*, relativamente a θ. Obtenha as distribuições *a posteriori* correspondentes. Para cada uma delas,

      - calcule a média, moda e mediana de θ *a posteriori*. Comente os resultados que obtiver.
      - indique o valor de $P(\theta > 0.6 \mid x)$.
      - calcule o factor de Bayes para testar $H_0 : \theta \geq 0.6$ vs. $H_1 : \theta < 0.6$. Faça os comentários que considerar adequados.

30. [6] Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias i.i.d. com distribuição de $Poisson(\theta)$, $\theta > 0$. Considere a distribuição $Gama(1, 1)$ como distribuição *a priori* para θ. Obtenha o intervalo de credibilidade a 90% de igual probabilidade e a região de credibilidade HPD a 90% para θ, considerando que possui uma amostra aleatória de dimensão $n = 10$ para a qual obteve $\sum_{i=1}^{n} x_i = 6$.

31. [7] Numa certa população, seja θ a proporção de indivíduos que têm uma determinada doença. Suponha que se pretende testar a hipótese $H_0 : \theta = 0.2$ contra a hipótese alternativa $H_1 : \theta = 0.3$. Informações anteriores mostram que $p(H_0) = 0.25$. Suponha que se observam n indivíduos verificando-se que x apresentam a doença. Calcule o factor de Bayes a favor de H_0. Para que valores de x é que se tem $p(H_0 \mid x) > p(H_1 \mid x)$?

[6] Trabalho 2005/2006
[7] Trabalho 2005/2006