Especificação e Estimação de Modelos Bayesianos

- Seguiremos o livro de Jim Albert “Bayesian Computation with R”
- Qual a ideia?
- Modelos com um parâmetro
- Modelos com múltiplos parâmetros

Que Proporção de Estudantes dormem 8 horas ou mais por noite?

- Pesquisador buscou na literatura estudos anteriores e descobre: um artigo em que a maioria dorme menos que 6 horas; outro artigo diz que 70 porcento dormem entre 5 e 6 horas, 28% entre 7 e 8, e 2% 9 horas.
- Fundamentado nessa informação o pesquisador acha que $p < 0.5$, provavelmente $p=0.3$.
- Ela faz uma pesquisa com 27 alunos onde 11 dormiram pelo menos 8 h na noite anterior.

Como juntar conhecimento anterior e dados novos?

$$L(p) \propto p^s (1-p)^f,\ 0 < p < 1.$$

The posterior density for p, by Bayes' rule, is obtained, up to a proportionality constant, by multiplying the prior density by the likelihood:

$$g(p|\text{data}) \propto g(p)L(p).$$

Como escolher a priori?

- Não tem uma resposta simples e geral.
- Para esse exemplo tentaremos três prioris:
  - Priori discreta
  - Priori contínua
  - Priori de histograma

Priori Discreta

- Crenças e pesos para valores específicos

$$.05, .15, .25, .35, .45, .55, .65, .75, .85, .95$$

are possible values for p. Based on her beliefs, she assigns these values the corresponding weights

$$1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0,$$

Ver R

Priori Discreta

- No exemplo tivemos 11 sucessos e 16 fracassos que leva à verossimilhança:

$$L(p) \propto p^{11}(1-p)^{16},\ 0 < p < 1.$$

- No pacote LearnBayes existe a função pdisc para estimar proporções com prioris discretas
- Ver R. Prob(p=0.25,0.35,0.45)= 94%

Priori Contínua (Distribuição Beta)

- Podemos alocar probabilidades para cada valor de 0 < p < 1 usando uma densidade
- Uma distribuição que aloca densidade nesse intervalo é a distribuição Beta:

$$g(p) \propto p^{a-1}(1-p)^{b-1}, \; 0 < p < 1,$$

- Mas, como escolher 'a' e 'b'? Uma possibilidade é fixar quantis. Por exemplo: Pr(p<0.3) = 0.5 e Pr(p<0.5) = 0.9. (ver R)

Priori Beta

- Com uma priori Beta(a,b) e dados com 's' sucessos e 'f' fracassos, a distribuição posteriori é:

$$g(p|\text{data}) \propto p^{a+s-1}(1-p)^{b+f-1},\ 0 < p < 1,$$

- Quando a distribuição priori e posteriori são da mesma família, chamamos a priori e a verossimilhança de **conjugadas**.
- Para nossa priori e dados temos a+s = 3.26+11 e b+f = 7.19+16. (ver R)

Priori Beta

- Estamos com a posteriori, e agora?
- Podemos usar os comandos 'pbeta' e 'qbeta' do R para responder perguntas específicas:
  - Qual a probabilidade que mais que 50% dos alunos dormem 8h ou mais? 1 – P(p<=0.5) = 1-pbeta(0.5,a+s,b+f)
  - Qual o intervalo de credibilidade da posteriori de 90%? qbeta(c(0.5,0.95),a+s,b+f)
  - Ver R para encontrar os valores.

Priori Beta

- Os cálculos da posteriori são exatos, pois conhecemos a distribuição posteriori. E se não conhecêssemos?
- Se tiver como gerar números aleatórios da distribuição posteriori podemos usar nossas estimativas Monte Carlo:
  - Pr(p>0.5) = fração de valores acima de 0.5
  - Intervalo de credibilidade de 90% é determinado pelos quantis 0.05 e 0.95 dos valores gerados.
  - Ver R

Priori de Histograma

- Método de força bruta:

- Choose a grid of values of p over an interval that covers the posterior density.
- Compute the product of the likelihood $L(p)$ and the prior $g(p)$ on the grid.
- Normalize by dividing each product by the sum of the products. In this step, we are approximating the posterior density by a discrete probability distribution on the grid.
- Using the R command `sample`, take a random sample with replacement from the discrete distribution.

Ver R

Previsão

- Com nosso conhecimento sobre 'p' (sendo esse conhecimento priori ou posteriori) podemos querer prever o número de sucessos ou fracassos numa amostra futura de 20 alunos
- A densidade preditiva (priori ou posteriori) é dada por:
- $$f(\tilde{y}) = \int f(\tilde{y}|p)g(p)dp.$$
- É prior se g(p) for a priori e posteriori se g(p) for a posteriori (aqui concentraremos em prioris)

Previsão (distribuições discretas)

- Para cada valor de 'p' o número de sucessos numa amostra de tamanho 20 segue um binomial:

$$f_B(y|n,p) = \binom{n}{y} p^y (1-p)^{n-y}, \ y = 0, ..., n.$$

- A probabilidade preditiva para y~ sucessos numa amostra futura de tamanho m é:

$$f(\tilde{y}) = \sum f_B(\tilde{y}|m, p_i) g(p_i).$$

- A função 'pdiscp' faz isso (ver R)

Previsão (contínua)

- Se modelar nossa crença sobre 'p' via uma priori Beta(a,b) a integral pode ser resolvida analiticamente para obter:

$$
\begin{aligned}
f(\tilde{y}) &= \int f_B(\tilde{y}|m,p)g(p)dp \\
&= \binom{m}{\tilde{y}} \frac{B(a+\tilde{y}, b+m-\tilde{y})}{B(a,b)}, \quad \tilde{y} = 0, ..., m,
\end{aligned}
$$

- A função 'pbetap' do pacote 'LearnBayes' faz esse cáculo. (ver R)

Previsão (simulação)

- Se ambos a distribuição de 'p' e o número de sucessos podem ser simulados podemos gerar uma amostra de distribuição preditiva pelos seguintes passos:
  - Amostrar alguns valores de 'p'
  - Para cada valor de 'p' amostrado, amostra um valor de 'Y', o número de sucessos
- O os valores de 'Y' vem da distribuição de previsão. (ver R)

Previsão (resumos da distribuição)

- Para resumir a distribuição preditiva por um intervalo de resultados que cobre 90% dos valores possíveis, podemos usar a função 'discint' do pacote 'LearnBayes'.
- A função recebe a distribuição dos 'y's e a probabilidade de cobertura desejada e retorna o intervalo e a probabilidade coberta pelo intervalo.
- Aqui temos duas fontes de incerteza. 'p' e o número de sucessos. (ver R)

Modelos com Um (1) Parâmetro

- Uma maneira de apostar em jogos de futebol americano é pelo número de pontos entre o vencedor e perdedor. Estudaremos a diferença entre o número que o 'bicheiro' oferece a uma chance de 50% e o número observado no jogo. Essa diferença, 'd' deve ter média zero, mas a variância é desconhecida. Vamos estimar essa variância. Sob a hipótese de normalidade temos:

$$L(\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\sum_{i=1}^{n} d_i^2/(2\sigma^2)\right\},\ \sigma^2 > 0.$$

Estimação da Variância

- Uma priori não informativa sobre a variância é que p(sigma) ~ 1/(sigma)^2
- Isso significa que o log(sigma) é uniformemente distribuída na reta (não sabemos a escala)
- Essa prior dá a seguinte posteriori:

$$g(\sigma^2|\mathrm{data}) \propto (\sigma^2)^{-n/2-1} \exp\{-v/(2\sigma^2)\}$$

- Onde 'v' é a soma dos d^2.
- O parâmetro P=1/(sigma)^2. P*v segue um Qui-quadrado com 'n' graus de liberdade. (ver R)

Taxa de Mortalidade (Transplante)

- Para dois hospitais 'n' (o número de transplantes), 'y' (o número de mortes) e 'e' (exposure = soma de 1/prob(morte)) são observados.
- Um modelo padrão assume que o número de mortes segue uma distribuição Poisson com média e*lambda. Queremos estimar lambda.
- Máxima verossimilhança dá lambda = y/e, mas isso pode ser ruim para y pequeno.
- Podemos usar informação priori.

Taxa de Mortalidade (Transplante)

- Se tiver informação sobre mortes (z) e exposição (o) de hospitais que achamos que tem uma taxa de mortalidade semelhante podemos criar uma distribuição para lambda.
  - Primeiro começa com uma priori não informativa, lambda ~ 1/lambda
  - Depois usa os dados dos 10 hospitais semelhantes
  - O conjugado de Poisson é Gamma

$$p(\lambda) \propto \lambda^{\sum_{j=1}^{10} z_j - 1} \exp\left(-(\sum_{j=1}^{10} o_j)\lambda\right)$$

- Esta distribuição serve de priori

Taxa de Mortalidade (Transplante)

- Se o número de mortes observados, y, segue uma Poisson(e*lambda) e a priori de lambda segue uma Gamma(alpha,beta), então a posteriori de lambda segue uma Gamma(alpha+y,beta+e)
- A distribuição preditiva (da priori) é:

$$f(y) = \frac{f(y|\lambda)g(\lambda)}{g(\lambda|y)},$$

where $f(y|\lambda)$ is the Poisson$(e\lambda)$ sampling density and $g(\lambda)$ and $g(\lambda|y)$ are, respectively, the prior and posterior densities of λ.

Taxa de Mortalidade (Transplante)

- Para verificar esses modelos, é importante comparar o resultado obtido de y com a distribuição de previsão de y.
- Se tiver no 'centro' da distribuição preditiva, então não temos razões de desacreditar o modelo.
- Se tiver nas cuadas da distribuição preditiva, então pode ser que nossa priori ou nosso modelo de amostra sejam errados.
- (Ver R)

Robustez Bayesiano

- As vezes várias prioris podem refletir nossas crenças
- Nesses casos é importante verificar que os resultados finais não são sensíveis à escolha da priori
- Exemplo: João tem um QI típico, theta, então achamos que a mediana é 100 e 90% dos valores estão entre 80 e 120.
- Ver R para a escolha de uma priori possível.

Robustez Bayesiano

- João faz quatro exames de QI. Assumindo que são independentes com média theta e desvio padrão 15, a média das quatro medições segue uma Normal(theta, sigma/sqrt(4))
- A posteriori é encontrada com precisão P=1/sigma^2 igual à soma das precisões da priori e dos dados.
- A média da posteriori é a média ponderada pelas precisões da priori e dos dados (ver próxima página)

Robustez Bayesiano

$$P_1 = P_D + P = 4/\sigma^2 + 1/\tau^2,$$

The posterior standard deviation is given by

$$\tau_1 = 1/\sqrt{P_1} = 1/(\sqrt{4/\sigma^2 + 1/\tau^2}).$$

The posterior mean of θ can be expressed as a weighted average of the sample mean and the prior mean where the weights are proportional to the precisions:

$$\mu_1 = \frac{\bar{y}P_D + \mu P}{P_D + P} = \frac{\bar{y}(4/\sigma^2) + \mu(1/\tau^2)}{4/\sigma^2 + 1/\tau^2}.$$

- Calculamos as estimativas para três médias diferentes, 110, 125 e 140. (ver R)

Robustez Bayesiano

- Uma densidade priori alternativa é a distribuição t de student com dois graus de liberdade (ver R para comparar com a Normal)

We perform the posterior calculations using the t prior for each of the possible sample results. Note that the posterior density of θ is given, up to a proportionality constant, by

$$g(\theta|\text{data}) \propto \phi(\bar{y}|\theta, \sigma/\sqrt{n}) g_T(\theta|v, \mu, \tau),$$

where $\phi(y|\theta, \sigma)$ is a normal density with mean θ and standard deviation σ, and $g_T(\mu|v, \mu, \tau)$ is a t density with median μ, scale parameter τ, and degrees of freedom v. Since this density does not have a convenient functional

Ver R

Misturas de Prioris Conjugadas

- Vimos que para os modelos de amostragem Normal, Poisson e Binomial, as prioris podem ser escolhidas para ter posterioris da mesma distribuição. (são conjugadas)
- Estes modelos podem parecer pouco flexíveis, mas nesses casos, misturas dessas prioris são também conjugadas.

Misturas de Prioris Conjugadas

- Considera uma moeda viciada num sentido desconhecido
- Podemos usar uma priori que mistura distribuições betas centradas em 0.3 e 0.7 com pesos de mistura iguais (gamma=0.5)
- $g(p) = \gamma g_1(p) + (1-\gamma) g_2(p)$
- A posteriori após observar os dados é

$$g(p|\text{data}) = \gamma(\text{data}) g_1(p|\text{data}) + (1-\gamma(\text{data})) g_2(p|\text{data})$$

- Com $\gamma(\text{data}) = \dfrac{\gamma f_1(s,f)}{\gamma f_1(s,f) + (1-\gamma) f_2(s,f)}$

Misturas de Prioris Conjugadas

- No coeficiente de mistura, f_i é a distribuição preditiva para caso i.

$$\gamma(\text{data}) = \frac{\gamma f_1(s,f)}{\gamma f_1(s,f) + (1-\gamma) f_2(s,f)}$$

- A função 'binomial.beta.mix' calcula a posteriori dado o número de sucessos e fracassos observados.
- Ver R

Teste Bayesiano da Moeda Viciada

- Um teste clássico de uma moeda viciada se fundamenta num p-valor calculada como 2*min(P(Y<=Yobs),P(Y>=Yobs)). No caso de 5 caras em 20 lançamentos sob H0 de p=0.5, o p-valor é 0.042, o que levaria a rejeitar a hipótese nula.
- Pode usar um modelo Bayesiano onde duas possibilidades existem
  - A moeda é justa (p=0.5) com probabilidade 0.5
  - A moeda não é justa (p != 0.5)

Teste Bayesiano da Moeda Viciada

- O segundo modelo pode ser especificado como um modelo beta(a,a) onde valores menores de a levam a densidades que cobrem mais do intervalo entre 0 e 1. A priori total fica:
- $g(p) = .5I(p = .5) + .5I(p \neq .5)g_1(p)$
- E a posteriori é dada por
- $g(p|y) = \lambda(y)I(p = .5) + (1 - \lambda(y))g_1(p|y)$
- Onde lambda(y) representa a probabilidade de H0 ser verdadeiro.

Teste Bayesiano da Moeda Viciada

- lambda(y) é dada por: $\lambda(y) = \dfrac{.5p(y|.5)}{.5p(y|.5) + .5m_1(y)}$
- Onde m1(y) é a densidade preditiva priori usando a distribuição beta(a,a), dada por:
- $m_1(y) = \dfrac{f(y|p)g_1(p)}{g_1(p|y)}$
- Ver R para o cálculo (note como o valor de 'a' não importa)

Teste Bayesiano da Moeda Viciada

- E se o teste fosse fundamentado em a probabilidade de ser menor que o valor observado ao invés de igual ao valor observado?

$$\lambda(y) = \frac{.5P_0(Y \leq 5)}{.5P_0(Y \leq 5) + .5P_1(Y \leq 5)}$$

- Ver R para o resultado.

Modelos com Múltiplos Parâmetros

- Dados Normais com média e variância desconhecida
- Considera os tempos para terminar a maratona de Nova Iorque para homens entre 20-29 anos
- Para uma amostra aleatória da normal e priori não informativa (P(sigma, mu) ~1/sigma^2) a posteriori é:

$$g(\mu, \sigma^2|y) \propto \frac{1}{(\sigma^2)^{n/2+1}} \exp\left(-\frac{1}{2\sigma^2}(S + n(\mu - \bar{y})^2)\right)$$

Modelo Normal

- the posterior of μ conditional on σ^2 is distributed as $N(\bar{y}, \sigma/\sqrt{n})$
- the marginal posterior of σ^2 is distributed as $S\chi^{-2}_{n-1}$, where χ^{-2}_{ν} denotes an inverse chi-square distribution with ν degrees of freedom

- Isso permite simulação do posteriori pelo marginal da variância primeiro seguida pela condicional da média
- Intervalos de credibilidade marginais podem ser calculados dos dados gerados da distribuição conjunta
- Ver R

Modelo Multinomial

- Em 1998 CBS fez um levantamento sobre intenções de votos nos EUA entre Bush e Dukakis. De 1447 adultos, y1=727 apoiaram Bush, y2=583 Dukakis e y3=137 outros ou indecisos.
- A fração que apoia Bush (theta1) é maior que a fração que apoio Dukakis (theta2)?
- Priori uniforme sobre as frações leva à posteriori Dirichlet: $g(\theta) = \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3}$
- Com theta1+theta2+theta3=1

Modelo Multinomial

- Para examinar (theta1-theta2) o caminho mais natural é simular da posteriori para os três thetas. Depois pode examinar a distribuição de theta1-theta2 ou contar a fração de vezes que theta1>theta2.
- Ver R

Modelo Multinomial

- Vamos estimar o número de votos eleitorais do Obama em 2008.
- Para tal usaremos um conjunto de pesquisas de intenção de voto contendo uma fração que apoia Obama e uma fração que apoia McCain para cada estado.
- O número de votos eleitorais para Obama é dado por

$$EV_O = \sum_{j=1}^{51} EV_j I(\theta_{Oj} > \theta_{Mj})$$

Modelo Multinomial

- Assumiremos que cada pesquisa entrevistou 500 pessoas (suposição conservador)
- E uma priori uniforme sobre theta_O e theta_M leva a posterioris independentes para cada estado com parâmetros

  (theta_O; theta_M; 1-theta_O – theta_M) = (500*f_O + 1; 500*f_M+1, 500*(1-f_O-f_M)+1)
- Com a distribuição posteriori, dá para simular
- Ver R

Comparação de Duas Proporções

- Observamos y1 ~ Bin(n1,p1) e y2 ~Bin(n2,p2)
- Queremos saber se p1>p2 ou p2>p1
- Se p1 e p2 forem independentes, então cada um pode ter suas densidades prioris e posterioris da Beta. O cálculo da probabilidade de p1>p2 é direto.
- Exploramos a possibilidade de p1 e p2 serem associados

Comparação de Duas Proporções

- Se p1=0.8 isso pode afetar nossa crença sobre p2 (fazer ele aparecer mais perto de 0.8, por exemplo)
- O modelo de Howard é: $\theta_1 = \log \frac{p_1}{1-p_1}, \theta_2 = \log \frac{p_2}{1-p_2}$
- $g(p_1, p_2) \propto e^{-(1/2)u^2} p_1^{\alpha-1}(1-p_1)^{\beta-1} p_2^{\gamma-1}(1-p_2)^{\delta-1}, 0 < p_1, p_2 < 1,$
- Onde $u = \frac{1}{\sigma}(\theta_1 - \theta_2)$
- Sigma controla quão próximas as proporções devem ser. Sigma pequena → próximas
- Alpha e beta: priori de p1
- Gamma e delta: priori de p2

Comparação de Duas Proporções

- Com dados binomiais a verossimilhança é:
- $L(p_1, p_2) \propto p_1^{y_1}(1-p_1)^{n_1-y_1} p_2^{y_2}(1-p_2)^{n_2-y_2}, 0 < p_1, p_2 < 1$
- E a posteriori segue a priori com parâmetros atualizados: $(\alpha + y_1, \beta + n_1 - y_1, \gamma + y_2, \delta + n_2 - y_2, \sigma)$
- Examinaremos a tabela a seguir:
- Ver R

Table 4.3. Pearson's example.

	Successes	Failures	Total
Sample 1	3	15	18
Sample 2	7	5	12
Totals	10	20	30