Inferência Bayesiana

Conceitos Básicos e Algumas Aplicações

Prof. Fábio N. Demarqui

2012/01

- Thomas Bayes (1702-1761) foi foi um matemático e pastor presbiteriano inglês.
- Em 1719 ingressou na Universidade de Edinburgh para estudar lógica e teologia.
- Mudou-se para Tunbridge Wells, Kent, por volta de 1734, onde permaneceu como ministro da Capela do Monte Sião até 1752.
- Thomas Bayes é conhecido por ter formulado o caso especial do teorema de Bayes (artigo publicado por um de seus pupilos após sua morte).
- Bayes foi eleito membro da Royal Society em 1742.

Figura: Thomas Bayes (1702-1761).

"O problema fundamental da estatística é a inferência. Dados são coletados e a partir deles desejamos fazer declarações (inferências) sobre uma ou mais características desconhecidas do mecanismo (ou processo) que deu origem aos dados observados."

- *A inferência estatística lida com o problema de tirar conclusões sobre quantidades não observadas a partir de dados numéricos (quantidades observadas).*
- *Tais quantidades não observadas podem ser de duas naturezas:*
    - *Quantidades que não são diretamente observáveis, tais como parâmetros que governam o processo hipotético que produz os dados observados;*
    - *Quantidades potencialmente observáveis, tais como observações futuras de um processo.*
- *As questões acima podem ser resolvidas utilizando-se tanto a inferência clássica (ou frequentista) quanto a inferência Bayesiana.*

Exemplo

- *O fabricante de um tipo de isolador elétrico quer conhecer o comportamento de seu produto funcionando a uma temperatura de* $200^{\circ}C$.
- *60 isoladores foram colocados sob teste sob a condição desejada até que 45 deles falhassem (censura do tipo II).*
- *O fabricante está interessado em saber:*
  - *O tempo médio (MTTF) e mediano de vida do isolador.*
  - *O percentual de falhas após 500 horas de uso.*

- O tempo de vida dos isoladores elétricos está sujeito a variações aleatórias, isto é, não é possível prevermos com exatidão qual será o tempo de vida de um determinado componente eletrônico.
- É razoável assumirmos, entretanto, que exista algum processo gerador dos dados (tempos de vida), possivelmente governado por um conjunto de parâmetros.
- Tal processo pode ser representado por uma distribuição de probabilidades.
- Podemos, então aceitar os dados observados como a realização de uma variável aleatória Y, ou um conjunto de v.a.'s $Y = (Y_1, ..., Y_n)$, com uma função de distribuição F_0 que representa a variabilidade ou incerteza da observação de Y.

- Evidentemente, a função de distribuição F_0 não é perfeitamente conhecida.
- No entanto, em geral, existe algum conhecimento inicial sobre a natureza do processo gerador dos dados que leva a propormos uma família de distribuições $\mathscr{F}$ a que pertence F_0, e que chamaremos de modelo estatístico.
- Formalmente, definimos um modelo estatístico da seguinte forma

$$\mathscr{F} = \{f(y|\theta) : \theta \in \Theta\}, \ y \in \mathscr{Y}, \quad (1)$$

em que $\mathscr{Y}$ corresponde ao espaço amostral associado ao experimento e Θ é chamado espaço paramétrico.
- Estamos interessados em encontrar o valor (fixo e desconhecido) $\theta_0 \in \Theta$ que nos leva a F_0.

- A inferência clássica baseia-se no princípio da repetitibilidade.
- Uma vez determinado o modelo estatístico, tanto a estimação de parâmetros quanto testes de hipóteses sobre os parâmetros são realizados levando-se em conta a variabilidade inerente à amostra observada.
- Em outras palavras, a idéia central da inferência clássica consiste em considerar-se a variabilidade que seria observada caso um mesmo experimento fosse repetido, sob as mesmas condições, um grande número de vezes.
- A teoria de verossimilhança desempenha papel de destaque na inferência clássica.

Exemplo

- *Suponha que estamos interessados em saber qual a proporção atual (digamos θ) de mulheres na Europa.*
- *Evidentemente, além de muito demorado, seria economicamente inviável contar o número total de mulheres.*
- *Ao invés disso, podemos tentar tirar nossas conclusões com base em uma amostra de n pessoas escolhidas aleatoriamente da população em estudo.*
- *Seja $X_i = 1$ se a i-ésima pessoa selecionada for uma mulher, e $X_i = 0$ caso contrário.*
- *É natural assumirmos que $X_1, ..., X_n$ representa uma amostra aleatória de $X \sim \text{Bernoulli}(\theta)$, tal que*

$$P(X_i = 1) = \theta \quad e \quad P(X_i = 0) = 1 - \theta. \tag{2}$$

- Um candidato "natural" para θ_0 neste caso seria $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Evidentemente, o valor de $\hat{\theta}$ está sujeito a variações da amostra, isto é, diferentes amostras observadas levarão a diferentes valores de $\hat{\theta}$.
- Observe que $\hat{\theta}$ é uma função das v.a.'s $X_1, \ldots, X_n$ ($\hat{\theta}$ é de fato uma estatística), logo faz sentido estudarmos a sua distribuição de probabilidades.
- Em inferência clássica, conclusões acerca do parâmetro de interesse θ_0 são feitas com base na informação da amostra observada, considerando-se as possíves variações dos valores observados quando da coleta de diferentes amostras.

- Defina $Y = \sum_{i=1}^{n} X_i$, então $Y \sim Bin(n, \theta)$. Então, nosso modelo estatístico assume a seguinte forma:

$$f(y|\theta) = \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}, \; y = 0, 1, 2, ..., n. \tag{3}$$

- Note que, para qualquer valor $\theta \in \Theta$, $f(y|\theta)$ é uma probabilidade!
- Como escolher então um valor $\hat{\theta} \in \Theta$ levando em conta apenas a informação de que dispomos, isto é, o valor observado y?

- Defina $Y = \sum_{i=1}^{n} X_i$, então $Y \sim Bin(n, \theta)$. Então, nosso modelo estatístico assume a seguinte forma:

$$f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \; y = 0, 1, 2, ..., n. \tag{3}$$

- Note que, para qualquer valor $\theta \in \Theta$, $f(y|\theta)$ é uma probabilidade!
- Como escolher então um valor $\hat{\theta} \in \Theta$ levando em conta apenas a informação de que dispomos, isto é, o valor observado y?
- Basta tomarmos o valor de θ que seja mais provável (verossímel) para a amostra observada!

- Matematicamente, o valor de $\hat{\theta}$ mais provável dada a amostra observada é dado por

$$\hat{\theta} = \sup_{\theta \in \Theta} L(\theta|y) = f(y|\theta). \tag{4}$$

- A função $L(\theta|y) = f(y|\theta)$ é chamada função de verossimilhança.
- Condicional nos dados observados, $L(\theta|y)$ deve ser vista como uma função de θ.
- Observe que é através de $L(\theta|y)$ que a informação dos dados observados é utilizada no processo inferencial.
- Pode ser mostrado que o valor $\hat{\theta}$ que maximixa $L(\theta|y)$ é $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Dizemos então que $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é o estimador de máxima verossimilhança de θ.

Exemplo

Vejamos um exemplo numérico...

- A inferência Bayesiana adota uma postura subjetivista através do uso explícito de probabilidades para quantificar o grau de incerteza acerca de quantidades de interesse não observadas.
- O objetivo da inferência Bayesiana consiste em combinar toda a informação subjetiva disponível referente a um problema, com a informação proveniente dos dados observados, através de declarações probabilísticas via teorema de Bayes.
- Etapas da análise Bayesiana:
  1. Especificação de um modelo probabilístico "completo": distribuição conjunta para todas as quantidades observadas e não observadas do problema em questão.
  2. Obtenção da distribuição a *posteriori*: distribuição condicional das quantidades não observadas dado os dados observados.
  3. Verificação da adequação do modelo ajustado aos dados.

- O que vem a ser uma probabilidade subjetiva?

- O que vem a ser uma probabilidade subjetiva?
- Qual time sairá vencedor no próximo confronto entre Cruzeiro e Atlético?

- O que vem a ser uma probabilidade subjetiva?
- Qual time sairá vencedor no próximo confronto entre Cruzeiro e Atlético?
- Obviamente, cruzeirenses e atleticanos têm opiniões bastante distintas quanto ao time que sairá vitorioso.

- O que vem a ser uma probabilidade subjetiva?
- Qual time sairá vencedor no próximo confronto entre Cruzeiro e Atlético?
- Obviamente, cruzeirenses e atleticanos têm opiniões bastante distintas quanto ao time que sairá vitorioso.
- Informalmente, podemos definir a probabilidade subjetiva como a crença que o observador do experimento possui na ocorrência do evento de interesse.

- O que vem a ser uma probabilidade subjetiva?
- Qual time sairá vencedor no próximo confronto entre Cruzeiro e Atlético?
- Obviamente, cruzeirenses e atleticanos têm opiniões bastante distintas quanto ao time que sairá vitorioso.
- Informalmente, podemos definir a probabilidade subjetiva como a crença que o observador do experimento possui na ocorrência do evento de interesse.
- Voltando à próxima partida entre Atlético e Cruzeiro, as chances de vitória que um torcedor atribui ao seu time do coração podem ser interpretadas como probabilidades subjetivas.

- O que vem a ser uma probabilidade subjetiva?
- Qual time sairá vencedor no próximo confronto entre Cruzeiro e Atlético?
- Obviamente, cruzeirenses e atleticanos têm opiniões bastante distintas quanto ao time que sairá vitorioso.
- Informalmente, podemos definir a probabilidade subjetiva como a crença que o observador do experimento possui na ocorrência do evento de interesse.
- Voltando à próxima partida entre Atlético e Cruzeiro, as chances de vitória que um torcedor atribui ao seu time do coração podem ser interpretadas como probabilidades subjetivas.
- Evidentemente, a probabilidade subjetiva deve ser coerente no sentido de satisfazer os axiomas de probabilidade...

- Suponha que tenhamos interesse em estimar um parâmetro θ, para o qual dispomos de alguma informação *a priori* que julgamos ser relevante.
- Tal informação pode ser representada probabilisticamente através da chamada distribuição *a priori* de θ, que denotaremos por $\pi(\theta)$.
- Assim como na abordagem frequentista, toda a informação proveniente dos dados observados é carregada pela função de verossimilhança $L(\theta|y) = f(y|\theta)$.
- A informação contida em $\pi(\theta)$ é, então, atualizada através da informação dos dados contida em $L(\theta|y) = f(y|\theta)$, via teorema de Bayes, levando à distribuição *a posteriori* de θ, dada por

$$\pi(\theta|y) = \frac{L(\theta|y)\pi(\theta)}{\int_{\Theta} L(\theta|y)\pi(\theta)d\theta}. \quad (5)$$

- A distribuição *a posteriori* de θ contém toda a informação de que dispomos para fazermos inferências sobre θ.
- É a partir dela que todas as estatísticas necessárias para descrevermos o comportamento de θ são extraídas.
- Assim, de posse da distribuição *a posteriori* de θ, podemos encontrar, por exemplo, o percentil de ordem $100(1-\alpha)\%$ de θ,

$$P_\alpha = \left\{ \theta' \in \Theta : \int_{-\infty}^{\theta'} \pi(\theta|y)d\theta \right\} = \alpha, \quad 0 < \alpha < 1. \tag{6}$$

- Estimativas pontuais para θ podem ser obtidas a partir de alguma medida de centralidade associada a distribuição *a posteriori* de θ . Três escolhas comuns são:
  - Média *a posteriori*,
  $$\hat{\theta} = E(\theta|y) = \int_{\Theta} \theta\pi(\theta|y)d\theta;$$
  - Mediana *a posteriori*,
  $$\hat{\theta} = P_{0.5};$$
  - Moda *a posteriori*,
  $$\hat{\theta} : \pi(\hat{\theta}|y) = \sup_{\theta} \pi(\theta|y).$$

- A quantidade

$$f(y) = \int_{\Theta} L(\theta|y)\pi(\theta)d\theta;$$

é chamada distribuição preditiva a *priori* de y.

- Predições podem ser feitas através da distribuição preditiva a *posteriori* de y,

$$f(\tilde{y}) = \int_{\Theta} f(\tilde{y}|\theta)\pi(\theta|y)d\theta;$$

- A distribuição *a posteriori* de θ também nos permite fazer declarações probabilísticas sobre θ. Por exemplo, podemos encontrar $\theta^{'}$ e $\theta^{''}$ a partir da solução de

$$\theta^{'} = \int_{-\infty}^{\theta^{'}} \pi(\theta|y)d\theta = \alpha/2 \text{ e } \theta^{''} = \int_{\theta''}^{\infty} \pi(\theta|y)d\theta = \alpha/2, \qquad (7)$$

tal que $P(\theta^{'} < \theta < \theta^{''}) = 1 - \alpha$.

- Desta forma, o conjunto $(\theta^{'}, \theta^{''})$ representa um intervalo de credibilidade de $100(1-\alpha)\%$ para θ.
- Alternativamente, podemos ainda obter o chamado intervalo HPD (highest posterior density), que corresponde à região $\Theta^{'} \subset \Theta$ de mais alta densidade de $\pi(\theta|y)$.

- Na escolha por uma família de distribuições a *priori* são desejáveis as propriedades:
    - Versatilidade para acomodar o maior número possível de crenças a *priori*.
    - Interpretabilidade para facilitar o processo de sumariação dos seus membros.
    - Simplicidade da derivação analítica das distribuições a *posteriori* e preditivas.
- Na prática tais propriedades desejaveis estão, em geral, em dessintonia.
- A simplicidade da operação Bayesiana poderá ficar garantida se impusermos algumas restrições sobre a família de distribuições a *priori* de θ, como veremos nos exemplos a seguir.

Exemplo

Retornando ao exemplo da proporção de mulheres na Europa, suponha que uma amostra de tamanho n foi coletada. Temos

- *Função de verossimilhança*

$$L(\theta|y) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y}, \tag{8}$$

- *Distribuição a priori:*

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\ 0<\theta<1 \tag{9}$$

que corresponde uma distribuição beta com parâmetros $\alpha > 0$ *e* $\beta > 0$, *(denotaremos* $\theta \sim Beta(\alpha;\beta)$*).*

- Se $\theta \sim Beta(\alpha;\beta)$, então

$$E(\theta)=\frac{\alpha}{\alpha+\beta}. \tag{10}$$

$$moda(\theta)=\frac{\alpha-1}{\alpha+\beta-2} \tag{11}$$

$$VAR(\theta)=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} \tag{12}$$

- A distribuição beta é uma distribuição bastante rica em formas...

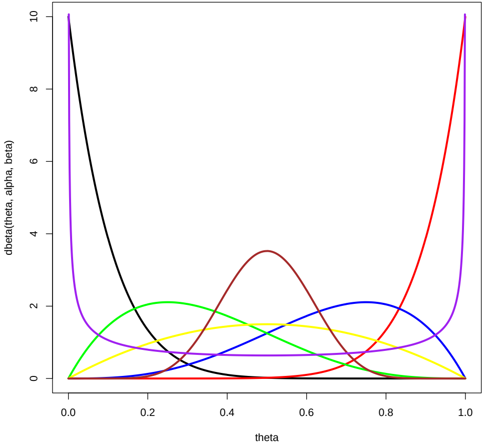

Figura: Distribuição *Beta*$(\alpha;\beta)$ segundo diferentes escolhas de α e β.

- Modelo probabilístico completo:

$$\begin{aligned} \pi(y,\theta) &= L(\theta|y)\pi(\theta) \\ &= \begin{pmatrix} n \\ y \end{pmatrix} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1} \end{aligned} \tag{13}$$

- Distribuição preditiva a *priori*:

$$\begin{aligned} f(y) &= \int_0^1 f(y|\theta)\pi(\theta)d\theta \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+n)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta+n)}. \end{aligned} \tag{14}$$

- Distribuição a *posteriori*:

$$\pi(\theta|y) = \frac{\Gamma(\alpha_* + \beta_*)}{\Gamma(\alpha_*)\Gamma(\beta_*)}\theta^{\alpha_*-1}(1-\theta)^{\beta_*-1} \tag{15}$$

  em que $\alpha_* = \alpha + y$ e $\beta_* = \beta + n - y$.

- Note que $\pi(\theta|y)$ pertence à mesma família de distribuições (família Beta) da distribuição $\pi(\theta)$.

- Dizemos então que a família de distribuições Beta é a família conjugada natural da família de distribuições Binomial (modelo estatístico assumido neste exemplo).

- De fato, a família de distribuições Beta corresponde à família conjugada natural das famílias de distribuições Bernoulli, Binomial e Geométrica.

- Voltando ao nosso exemplo, a título de ilustração, vamos consideradas as seguintes distribuições *priori* para θ:

Tabela: Distribuições a *priori* de θ.

$\pi(\theta)$	$E(\theta)$	$moda(\theta)$	$VAR(\theta)$
$Beta(1,1)$	0.5	0.5	0.0883
$Beta(10,10)$	0.5	0.5	0.0119
$Beta(1,10)$	0.0909	-	0.0069

- Graficamente, temos...

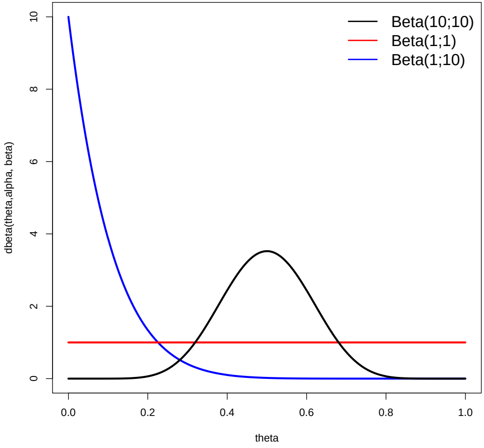

Figura: Distribuição a *priori* de θ.

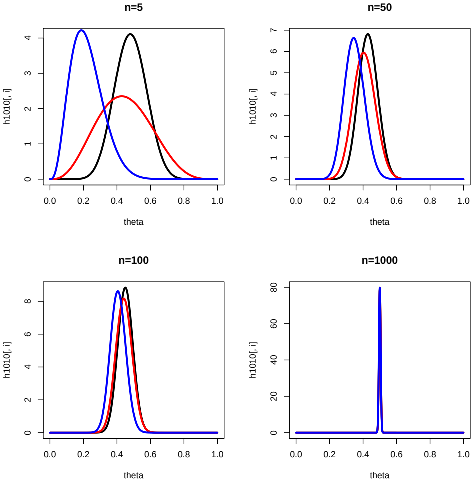

Figura: Distribuição a *posteriori* de θ.

Exemplo

- *Vamos retornar agora ao exemplo do tempo de vida dos isoladores elétricos.*
- *Suponha que temos motivos para acreditar que a distribuição exponencial*

$$f(y|\theta) = \theta e^{-\theta y}, \; \theta > 0, \tag{16}$$

  corresponde a um modelo estatístico adequado para o problema.
- *A família Gama de distribuições corresponde a família conjugada natural da distribuição exponencial.*
- *Assim, assumiremos que*

$$\pi(\theta) = \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1} e^{-\gamma_0 \theta}, \; \alpha_0 > 0, \gamma_0 > 0. \tag{17}$$

- Se $\theta \sim Gama(\alpha_0, \gamma_0)$, então:

$$E(\theta) = \frac{\alpha_0}{\gamma_0} \tag{18}$$

e

$$V(\theta) = \frac{\alpha_0}{\gamma_0^2}. \tag{19}$$

- É fácil mostrar que $\alpha_0 = \dfrac{E(\theta)^2}{V(\theta)}$ e $\gamma_0 = \dfrac{E(\theta)}{V(\theta)}$.
- Logo, se a informação subjetiva disponível puder ser expressada em termos de $E(\theta)$ e $V(\theta)$, é possível especificarmos os valores de α_0 e γ_0 de modo a refletir nosso conhecimento a *priori* sobre o problema.

- Função de verossimilhança:

$$L(\theta|y) = f(y_1, ..., y_n|\theta) = \theta^{\sum_{i=1}^{n} \delta_i} \exp\left\{-\theta \sum_{i=1}^{n} y_i\right\} \tag{20}$$

em que $y = (y_1, ..., y_n)$.

- Modelo probabilístico completo:

$$\begin{aligned} \pi(y, \theta) &\propto L(\theta|y)\pi(\theta) \\ &\propto \theta^{\alpha_0 + \sum_{i=1}^{n} \delta_i - 1} \exp\left\{-\theta \left[\gamma_0 + \sum_{i=1}^{n} y_i\right]\right\}, \end{aligned} \tag{21}$$

que corresponde ao núcleo da distribuição gama, isto é,

$$(\theta|y) \sim Gama\,(\alpha; \gamma)\,,$$

em que $\alpha = \alpha_0 + \sum_{i=1}^{n} \delta_i$ e $\gamma = \gamma_0 + \sum_{i=1}^{n} y_i$.

- Distribuição preditiva a *priori*:

$$
\begin{aligned}
f(y) &= \int_0^\infty f(y|\theta)\pi(\theta)d\theta \\
&= \frac{\gamma_0^{\alpha_0}}{\Gamma(\alpha_0)} \frac{\Gamma(\alpha_0 + \sum_{i=1}^n \delta_i)}{\left[\gamma_0 + \sum_{i=1}^n y_i\right]^{(\alpha_0 + n\sum_{i=1}^n \delta_i)}}.
\end{aligned} \tag{22}
$$

- Distribuição preditiva a *posteriori*:

$$
\begin{aligned}
f(y_{n+1}|y) &= \int_0^\infty f(y_{n+1}|\theta)\pi(\theta|y)d\theta \\
&= \frac{\gamma^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{[\gamma + y_{n+1}]^{(\alpha+1)}}.
\end{aligned} \tag{23}
$$

- Função de confiabilidade a *posteriori*:

$$
\begin{aligned}
R(t|y) + P(Y_{n+1} > t|y) &= \int_0^\infty e^{-\theta t}\pi(\theta|y)d\theta \\
&= \left(\frac{\gamma}{\gamma + t}\right)^{(\alpha)}
\end{aligned} \tag{24}
$$

- Dizemos que uma *priori* $\pi(\theta)$ é imprópria se

$$\int_\Theta \pi(\theta)d\theta = \infty. \tag{25}$$

- *Priori* impróprias são frequentemente utilizadas em inferência Bayesiana $\rightarrow$ *priori* **não informativas**.
- Observações:
  - *Priori* impróprias podem levar a *posteriori* impróprias.
  - Não podemos fazer inferências com base em uma *posteriori* imprópria.
  - Para verificarmos se uma *posteriori* é própria, basta mostrarmos que

$$\int_\Theta f(y|\theta)\pi(\theta)d\theta < \infty, \tag{26}$$

    o que nem sempre é uma tarefa fácil.
  - Se $\pi(\theta)$ é própria, então $\pi(\theta|y)$ também é própria! (Por que???)

- Seja $f(y|\theta)$. Então a matriz de informação de Fisher esperada é dada por

$$I(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2} log\ f(y|\theta)\right] \qquad (27)$$

- A *priori* de Jeffreys é definida da seguinte forma

$$\pi(\theta) \propto |I(\theta)|^{1/2}, \qquad (28)$$

- A *priori* de Jeffreys é invariante com respeito a transformações.
- Seja $\phi = h(\theta)$. Então,

$$\pi(\phi) = \pi(\theta)\left\|\frac{d\theta}{d\phi}\right\|, \qquad (29)$$

em que $\theta = h^{-1}(\phi)$

- Dizemos que uma *priori* $\pi(\theta)$ é flat (ou vaga) se

$$V(\theta) \to \infty \tag{30}$$

- *Priori* flat são bastante utilizadas quando temos interesse em usar uma *priori* que seja não informativa.
- Observações:
  - A *posteriori* resultante é sempre própria.
  - Os softwares WinBUGS e OpenBUGS não aceitam *priori* impróprias.

- Em geral, uma expressão fechada para a distribuição a *posteriori*

$$\pi(\theta|y) = \frac{L(\theta|y)\pi(\theta)}{\int_{\Theta} L(\theta|y)\pi(\theta)d\theta} \tag{31}$$

  não pode ser obtida, pois a integral que aparece no denominador em (31) não possui solução analítica.
- Em situações como esta, a distribuição *a posteriori* pode ser obtida através de métodos numéricos.
- Neste cenário, métodos de simulação via Monte Carlo via Cadeias de Markov (MCMC) tem sido amplamente utilizados, com destaque para:
  - Amostrador de Gibbs (Gibbs Sampler);
  - Algoritmo de Metropolis-Hastings;

- Para a implementação de métodos MCMC precisamos conhecer apenas o núcleo da distribuição a *posteriori* de θ, isto é,

$$\pi(\theta|y) \propto L(\theta|y)\pi(\theta). \tag{32}$$

- A idéia básica é que se conhecemos o núcleo de $\pi(\theta|y)$, então é possível gerarmos amostras dessa distribuição usando teoria de cadeias de Markov.
- Com uma amostra da distribuição a *posteriori* de θ, características de $\pi(\theta|y)$ podem ser estudadas empiricamente através de técnicas descritivas.

- Um dos métodos mais comuns de seleção de modelos se dá através da comparação das probabilidades a *posteriori* dos modelos sob investigação.
- Seja m um modelo pertencente aos espaço de modelos $\mathscr{M}$.
- A probabilidade a *posteriori* do modelo m é dada por

$$\pi(m|y) = \frac{\pi(D|m)\pi(m)}{\sum_{m \in \mathscr{M}} \pi(D|m)\pi(m)}, \tag{33}$$

em que

$$\pi(D|m) = \int L(\theta^m|y)\pi(\theta^m)d\theta^m, \tag{34}$$

e $\pi(m)$ denota a probabilidade a *priori* do modelo m.

- Sejam $\mathscr{M}_1$ e $\mathscr{M}_2$ dois modelos a serem comparados.
- O Fator de Bayes (BF) é definido como

$$\begin{aligned} BF(\mathscr{M}_1, \mathscr{M}_1) &= \frac{\pi(\mathscr{M}_1|y)/\pi(\mathscr{M}_1|y)}{\pi(\mathscr{M}_1)/\pi(\mathscr{M}_1)} \\ &= \frac{\pi(y|\mathscr{M}_1)}{\pi(y|\mathscr{M}_2)}, \end{aligned} \tag{35}$$

em que

$$\pi(y|\mathscr{M}_k) = \int L(\theta_k|y)\pi(\theta_k)d\theta_k, \tag{36}$$

$k = 1, 2$.

- Evidentemente, o modelo com maior probabilidade a *posteriori* é preferível.

- Escala de evidência em favor de $\mathscr{M}_1$ proposta por Kass & Raftery (1995):

Tabela: Escala de evidência em favor de $\mathscr{M}_1$

1-3	não significativa
3-30	positiva
30-150	forte
> 150	mutio forte

- A deviance é definida da seguinte forma

$$D(y,\theta) = -2 \log f(y|\theta). \tag{37}$$

- O DIC é definido como

$$DIC = 2\hat{D}_{avg}(y) - D_{\hat{\theta}}(y) \tag{38}$$

em que

$$\hat{D}_{avg}(y) = \frac{1}{N}\sum_{l=1}^{N} D(y,\theta^{l}) \tag{39}$$

e

$$D_{\hat{\theta}}(y) = D(y,\hat{\theta}). \tag{40}$$

- A quantidade

$$p_D = \hat{D}_{avg}(y) - D_{\hat{\theta}}(y) \tag{41}$$

é frequentemente utilizada como uma medida do número efetivo de parâmetros no modelo Bayesiano.

- Modelos que apresentam DIC com menor valor são preferíveis.
- Problemas:
  i) p_D pode ser negativo em alguns casos.
  ii) não é recomendado para modelos com vários níveis de hieraquia.

- A quantidade

$$CPO_i = f(y_i|y_{-i}) = \int f(y_i|\theta)\pi(\theta|y_{-i})d\theta \tag{42}$$

é chamada conditional predictive ordinate.

- Uma medida bastante utilizada para a comparação de modelos em inferência Bayesiana é

$$LPML = \sum_{i=1}^{n} \log(CPO_i). \tag{43}$$

- Modelos com maior valor de LPML são preferíveis.
- Pode ser mostrado que

$$\widehat{CPO}_i \approx M \left\{ \sum_{l=1}^{M} [f(y_i|\theta^l)]^{-1} \right\}^{-1}, \quad l = 1, \ldots, M. \tag{44}$$

Inferência Clássica:

- baseia-se no princípio da repetibilidade;
- o conceito de aleatoriedade está associado a variabilidade verificada através da coleta de diferentes amostras;
- em geral, depende de resultados assintóticos.

Inferência Bayesiana:

- postura subjetivista;
- uso explícito da probabilidade para quantificar a incerteza;
- abribui-se probabilidades aos possíveis valores que o(s) parâmetro(s) pode(m) assumir;
- não depdende de resultados assintóticos.
- opiniões subjetivas coerentes $\Rightarrow$ axiomas de probabilidade são satisfeitos.

Vantagens:

- Permite a inclusão de informação subjetiva relevante durante o processo inferencial.
- Fornece maior flexibilidade na modelagem dos dados, especialmente para conjuntos de dados com estruturas mais complexas.

Desvantagens:

- Salvo algumas situações envolvento problemas relativamente simples, exige um custo computacional alto.
- A implementação de métodos MCMC pode não ser trivial, requerendo certo grau de experiência por parte do analista.