GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística

Universidade Federal Fluminense

Aula 08

Modelo Linear Normal

- Considere o seguinte modelo:

$$E(Y_i) = \mu_i = \mathbf{x}_i'\boldsymbol{\beta}, \quad Y_i \sim \mathcal{N}(\mu_i, \sigma^2), \quad i = 1, \ldots, n,$$

onde $Y_1, \ldots, Y_n$ são variáveis aleatórias independentes.

- A função de ligação é a identidade:

$$g(\mu_i) = \mu_i$$

- Esse modelo pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Assumindo que $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, para $i = 1, \ldots, n$.

Estimação no Modelo Linear Normal

- O estimador de máxima verossimilhança e o estimador de mínimos quadrados de $\boldsymbol{\beta}$ coincidem e são dados por:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y},$$

  onde $\mathbf{X}'\mathbf{X}$ é não singular.

- O estimador $\hat{\boldsymbol{\beta}}$ é não viciado e possui matriz de covariâncias dada por:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad V(\hat{\boldsymbol{\beta}}) = E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \boldsymbol{\mathcal{I}}^{-1}$$

- No contexto de MLG, σ^2 é tratado como parâmetro de ruído e seu estimador é não viciado:

$$\hat{\sigma}^2 = \frac{\text{SQRes}}{n-p} = \frac{1}{n-p}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \frac{1}{n-p}\left(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}\right)$$

- Esse estimador pode ser utilizado para estimar $\boldsymbol{\mathcal{I}}$ e, consequentemente, para realizar inferência sobre $\hat{\boldsymbol{\beta}}$.

Deviance no Modelo Linear Normal

- A deviance para o modelo linear normal, do ponto de vista matricial, é dada por:

$$D = \frac{1}{\sigma^2}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

- Expandindo:

$$D = \frac{1}{\sigma^2}\left[\mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\right]$$

- Como $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, obtemos:

$$D = \frac{1}{\sigma^2}\left[\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}\right].$$

- Note que a deviance depende do parâmetro σ^2. Neste caso, não é possível utilizá-la diretamente para comparar modelos, como é feito nos outros modelos da família exponencial.

- Para comparar dois modelos, considere a hipótese nula H_0 e uma hipótese alternativa mais geral H_1:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}, \quad H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},$$

  onde $q < p < n$.

- Sejam $\mathbf{X}_0$ e $\mathbf{X}_1$ as correspondentes matrizes de variáveis explicativas, $\hat{\boldsymbol{\beta}}_0$ e $\hat{\boldsymbol{\beta}}_1$ os EMVs, e D_0 e D_1 as deviances.

Teste de Hipóteses com Deviance

- Testamos H_0 versus H_1 usando:

$$\Delta D = D_0 - D_1 = \frac{1}{\sigma^2}\left[\left(\mathbf{y}'\mathbf{y} - \hat{\beta}_0'\mathbf{X}_0'\mathbf{y}\right) - \left(\mathbf{y}'\mathbf{y} - \hat{\beta}_1'\mathbf{X}_1'\mathbf{y}\right)\right]$$

$$= \frac{1}{\sigma^2}\left(\hat{\beta}_1'\mathbf{X}_1'\mathbf{y} - \hat{\beta}_0'\mathbf{X}_0'\mathbf{y}\right).$$

- Como o modelo sob H_1 é mais geral, ele tende a ajustar melhor os dados. Assim, assume-se que D_1 segue uma distribuição central:

$$D_1 \sim \chi^2_{n-p}$$

- Por outro lado, se H_0 não for verdadeira, D_0 pode seguir uma distribuição não-central:

$$D_0 \sim \chi^2_{n-q,\nu}$$

- Nesse caso, ΔD segue uma distribuição não-central:

$$\Delta D \sim \chi^2_{p-q,\nu}$$

Teste de Hipóteses via Deviance

- Sob H_0 verdadeira, a estatística

$$F = \frac{D_0 - D_1}{p-q} \Big/ \frac{D_1}{n-p} = \frac{\hat{\beta}_1' \mathbf{X}_1' \mathbf{y} - \hat{\beta}_0' \mathbf{X}_0' \mathbf{y}}{p-q} \Big/ \frac{\mathbf{y}'\mathbf{y} - \hat{\beta}_1' \mathbf{X}_1' \mathbf{y}}{n-p}$$

  segue uma distribuição central

$$F \sim F_{p-q,\, n-p}$$

- Se o valor de F for grande em relação à distribuição $F_{p-q,\, n-p}$, há evidências contra H_0.

- Esse teste de hipóteses pode ser resumido na tabela ANOVA, relacionando soma de quadrados, graus de liberdade e estatísticas F.

Resíduos no Modelo Linear Normal

- Para o modelo linear normal, os resíduos são definidos como:

$$e_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}} = y_i - \hat{\mu}_i,$$

onde $\hat{\mu}_i$ é o valor ajustado pelo modelo.

- A matriz de covariâncias do vetor de resíduos $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ é:

$$V(\mathbf{e}) = (\mathbf{I}_n - \mathbf{H})\, V(\mathbf{Y})\, (\mathbf{I}_n - \mathbf{H})' = \sigma^2(\mathbf{I}_n - \mathbf{H})$$

onde $\mathbf{I}_n$ é a matriz identidade $n \times n$ e

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

é chamada de matriz de projeção ou "chapéu".

Resíduos Padronizados e Diagnóstico do Modelo

- Os resíduos padronizados são definidos como:

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}},$$

onde h_{ii} é o i-ésimo elemento da diagonal da matriz de projeção $\mathbf{H}$, e $\hat{\sigma}^2$ é o estimador de σ^2.

- Os resíduos são usados para verificar a adequabilidade do modelo ajustado.

- Ferramentas de diagnóstico permitem avaliar:
  - Linearidade das relações entre variáveis.
  - Independência e normalidade dos erros.
  - Possíveis associações com variáveis explicativas que não foram incluídas no modelo.

Outros Diagnósticos no Modelo Linear

- Além dos resíduos, existem outros métodos para avaliar a adequação do modelo e identificar observações não usuais ou influentes.

- Um outlier é uma observação que não é bem ajustada pelo modelo, enquanto uma observação influente é aquela que exerce um efeito relativamente grande na inferência do modelo.

- O valor h_{ii} é chamado de ponto de alavancagem da i-ésima observação.

- Uma observação com alta alavancagem pode ter grande impacto no ajuste do modelo.

Diagnóstico de Observações Influentes

- Valores de h_{ii} maiores do que $2p/n$ ou $3p/n$ podem ser motivo de preocupação.

- Medidas que combinam resíduos padronizados e alavancagem incluem:

$$\text{DFITS}_i = \frac{r_i\sqrt{h_{ii}}}{\sqrt{1-h_{ii}}},$$

$$\text{Distância de Cook: } D_i = \frac{r_i^2 h_{ii}}{p(1-h_{ii})}.$$

- Valores grandes dessas estatísticas indicam que a i-ésima observação é influente.

Exemplo: Regressão Linear Múltipla

- O exemplo considera um modelo de regressão linear múltipla com todas as variáveis explicativas contínuas.

- A variável resposta é a porcentagem de calorias totais obtidas a partir de carboidratos complexos para 20 homens diabéticos dependentes de insulina, que seguiram uma dieta rica em carboidratos durante seis meses.

- Considerou-se que o cumprimento do regime estava relacionado com:
  - Idade (em anos),
  - Peso corporal (em relação ao peso ideal para a altura),
  - Outros componentes, como a porcentagem de calorias provenientes de proteínas.

Exemplo: Regressão Linear Múltipla

Carbohydrate y	Age x_1	Weight x_2	Protein x_3
33	33	100	14
40	47	92	15
37	49	135	18
27	35	144	12
30	46	140	15
43	52	101	15
34	62	95	14
48	23	101	17
30	32	98	15
38	42	105	14
50	31	108	17
51	61	85	19
30	63	130	19
36	40	127	20
41	50	109	15
42	64	107	16
46	56	117	18
24	61	100	13
35	48	118	18
37	28	102	14

Tabela: Carboidrato, idade, peso relativo e proteína para 20 homens diabéticos dependentes de insulina.

Exemplo: Regressão Linear Múltipla

- Considere o seguinte modelo (Modelo 1):

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}, \quad Y_i \sim N(\mu_i, \sigma^2)$$

onde o carboidrato y está linearmente relacionado com a idade x_1, o peso relativo x_2 e a proteína x_3, $i = 1, \ldots, n$.

- Nesse caso, podemos escrever em forma matricial:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

- Note que o modelo proposto inclui um intercepto β_0.

Exemplo: Regressão Linear Múltipla

- A solução da equação normal $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ é

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 36{,}9601 \\ -0{,}1137 \\ -0{,}2280 \\ 1{,}9577 \end{bmatrix}$$

- Usando os resultados $\mathbf{y}'\mathbf{y} = 29.368$ e $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 28.799{,}97$, obtemos uma estimativa para a variância:

$$\hat{\sigma}^2 = \frac{1}{n-p}\left(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}\right) = 35{,}5.$$

- Os erros padrão para os elementos de $\hat{\boldsymbol{\beta}}$ são obtidos a partir do cálculo da inversa da matriz de informação de Fisher $\boldsymbol{\mathcal{I}}^{-1} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$.

- No nosso exemplo, temos

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4{,}8158 & -0{,}0113 & -0{,}0188 & -0{,}1362 \\ -0{,}0113 & 0{,}0003 & 0{,}0000 & -0{,}0004 \\ -0{,}0188 & 0{,}0000 & 0{,}0002 & -0{,}0002 \\ -0{,}1362 & -0{,}0004 & -0{,}0002 & 0{,}0114 \end{bmatrix}$$

Exemplo: Regressão Linear Múltipla

Termo	Estimativa $\hat{\beta}_j$	Erro padrão
Constante	36,960	13,071
Idade	-0,114	0,109
Peso	-0,228	0,083
Proteína	1,958	0,635

Exemplo: Regressão Linear Múltipla

- Para ilustrar o uso da deviance, vamos testar a hipótese H_0 de que a resposta não depende da idade, isto é, $\beta_1 = 0$.

- O modelo correspondente (Modelo 0) é dado por:

$$E(Y_i) = \beta_0 + \beta_2 x_{i2} + \beta_3 x_{i3}$$

- Para esse modelo, obtemos a seguinte estimativa dos coeficientes:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 33{,}130 \\ -0{,}222 \\ 1{,}824 \end{bmatrix}$$

- E $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 28.761{,}89$,

Exemplo: Regressão Linear Múltipla

- O valor da estatística F é:

$$F = \frac{D_0 - D_1}{p - q} \Big/ \frac{D_1}{n - p} = \frac{\hat{\beta}_1'\mathbf{X}_1'\mathbf{y} - \hat{\beta}_0'\mathbf{X}_0'\mathbf{y}}{p - q} \Big/ \frac{\mathbf{y}'\mathbf{y} - \hat{\beta}_1'\mathbf{X}_1'\mathbf{y}}{n - p}$$
$$= \frac{28.799{,}97 - 28.761{,}89}{4 - 3} \Big/ \frac{29.368 - 28.799{,}97}{20 - 4} = \frac{38{,}36}{35{,}48}$$
$$\approx 1{,}08$$

- Esse valor não é significativo quando comparado com a distribuição $F_{1,16}$.

- Conclusão: os dados não fornecem evidências contra H_0, ou seja, a resposta parece não estar relacionada com a idade.

```
> anova(fit0,fit1)
Analysis of Variance Table

Model 1: Carbo ~ Peso + Proteina
Model 2: Carbo ~ Age + Peso + Proteina
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     17 606.02
2     16 567.66  1    38.359 1.0812 0.3139
```

Multicolinearidade

- A multicolinearidade ocorre quando temos variáveis explicativas altamente correlacionadas entre si, trazendo consequências indesejadas.

- Algumas consequências da multicolinearidade:

  - As colunas da matriz $\mathbf{X}$ podem ser linearmente dependentes, levando à singularidade de $\mathbf{X}'\mathbf{X}$ e comprometendo a equação de estimação:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

  - A solução $\hat{\boldsymbol{\beta}}$ fica instável, e mudanças pequenas nos dados podem causar grandes alterações nos valores estimados.
  - Os elementos de $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ serão grandes e, consequentemente, as variâncias e covariâncias dos elementos de $\hat{\boldsymbol{\beta}}$ também serão elevadas.
  - A seleção de modelos e a escolha do melhor subconjunto de variáveis explicativas podem se tornar difíceis devido à instabilidade das estimativas.

Multicolinearidade

- A multicolinearidade pode ser detectada pelo cálculo do Fator de Inflação da Variância (VIF).

- O VIF para a j-ésima variável é dado por:

$$VIF_j = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o coeficiente de determinação obtido pela regressão da j-ésima variável explicativa *versus* todas as demais variáveis explicativas.

- Interpretação:
  - Se a j-ésima variável não é correlacionada com as demais, então $VIF_j = 1$.
  - O VIF_j cresce conforme a correlação com outras variáveis aumenta.
  - Valores de $VIF_j > 5$ devem ser considerados preocupantes.

Aprendizagem Bayesiana

- A aprendizagem Bayesiana se inicia com a formulação numérica das crenças a priori sobre θ e da informação vinda dos dados y.

- Isso é feito a partir das distribuições de θ e y:

  1. Distribuição a priori: descreve nossas crenças sobre θ antes de observar os dados, $p(\theta)$.
  2. Verossimilhança: descreve como os dados y se comportam dado θ, $p(y \mid \theta)$.
  3. Distribuição a posteriori: após observar y, nossas crenças sobre θ são atualizadas para $p(\theta|y)$.

- A regra de atualização é a fórmula de Bayes:

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)}$$

onde $p(y) = \int p(y \mid \theta) p(\theta) d\theta$ é a distribuição preditiva de y.

Modelo Normal: Média e Variância Desconhecidas

- Suponha que

$$Y_1, \ldots, Y_n \mid \Theta \overset{iid}{\sim} N(\theta, \sigma^2)$$

- A função de densidade conjunta é:

$$p(y_1, \ldots, y_n \mid \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2 \right\}$$

- Considere a priori:

$$\theta \sim N(\mu_0, \tau^2), \quad \sigma^2 \sim \text{Gama-Inv}(a, b)$$

- A posteriori conjunta é:

$$\begin{aligned} p(\theta, \sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2 \right\} \\ \times \exp\left\{ -\frac{1}{2\tau^2} (\theta - \mu_0)^2 \right\} \\ \times (\sigma^2)^{-a-1} \exp\left\{ -\frac{b}{\sigma^2} \right\} \end{aligned}$$

Modelo Normal: Condicionais Completas

- Condicional completa de θ:

$$p(\theta \mid \sigma^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}\exp\left\{-\frac{1}{2\tau^2}(\theta-\mu_0)^2\right\}$$

- Condicional completa de σ^2:

$$p(\sigma^2 \mid \theta, \mathbf{y}) \propto (\sigma^2)^{-(n/2+a)-1}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\theta)^2+2b\right]\right\}$$

- Logo, temos:

$$(\theta \mid \sigma^2, \mathbf{y}) \sim N(\mu_1, \sigma_1^2), \quad (\sigma^2 \mid \theta, \mathbf{y}) \sim \text{Gama-Inv}(a_1, b_1)$$

Algoritmo de Gibbs: Ideia Básica

- As distribuições $p(\theta \mid \sigma^2, \mathbf{y})$ e $p(\sigma^2 \mid \theta, \mathbf{y})$ são conhecidas como distribuições condicionais completas.

- Recebem esse nome porque são distribuições condicionais de cada parâmetro dado todos os outros.

- A ideia do algoritmo de Gibbs é gerar valores dessas condicionais completas (fáceis de simular) ao invés de gerar diretamente a posteriori conjunta (difícil de simular).

- Procedemos iterativamente:

$$(\theta^{(t)}, \sigma^{2(t)}) \sim \begin{cases} \theta^{(t)} \sim p(\theta \mid \sigma^{2(t-1)}, \mathbf{y}) \\ \sigma^{2(t)} \sim p(\sigma^2 \mid \theta^{(t)}, \mathbf{y}) \end{cases}$$

até convergência para a posteriori conjunta $p(\theta, \sigma^2 \mid \mathbf{y})$.

Exemplo: Modelo Normal

- Considere o modelo:

$$y \mid \mu, \phi \sim N(\mu, 1/\phi)$$

onde μ e ϕ são desconhecidos, sendo $\phi = 1/\sigma^2$ a precisão.

- Suponha distribuições a priori:

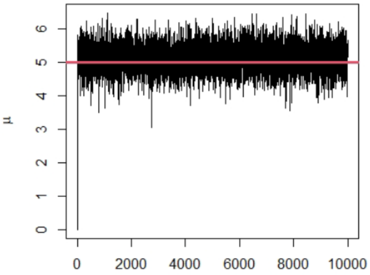
$$\mu \sim N(0, \tau_0^2), \quad \phi \sim \text{Gama}(a_0/2, b_0/2)$$

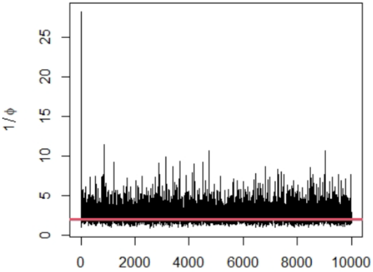
- As distribuições condicionais completas são:

$$\phi \mid \mu, y \sim \text{Gama}\left(\frac{n + a_0}{2}, \frac{\sum_{i=1}^{n}(y_i - \mu)^2 + b_0}{2}\right)$$

$$\mu \mid \phi, y \sim N(\mu_1, \tau_1^2), \quad \mu_1 = \tau_1^2 n\bar{y}\phi, \quad \tau_1^2 = \frac{1}{n\phi + \tau_0^{-2}}$$

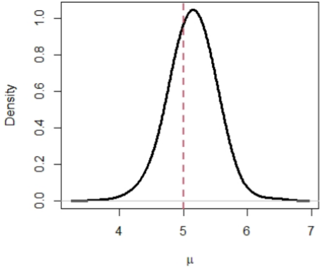
```r
## mu.true = 5 e sig2.true = 2
## dados simulados
y = c(4.05,5.13,6.93,4.85,5.72,5.50,
5.21,5.17,9.23,3.32,4.33,4.20,5.55,
7.96,3.49,4.71,6.12,4.10,2.82,5.85)

## valores iniciais
mu = 0; phi = 1

## dados
y.bar = mean(y); n = length(y)

## priori
a0 = b0 = 0.01
tau20 = 10

M = 10000

for(k in 2:M){
  s1 = sum((y-mu[k-1])^2)
  phi[k] = rgamma(1,(n+a0)/2,(s1+b0)/2)
  v21 = 1/(n*phi[k]+1/tau20)
  mu1 = v21*n*y.bar*phi[k]
  mu[k] = rnorm(1,mu1,sqrt(v21))
}
```

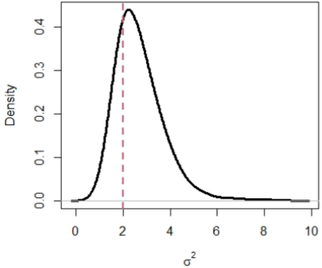




Exemplo: Modelo Normal

```
mu.sample = mu[seq(1000,M,by=9)]

plot(density(mu.sample,adjust=2),main="",
     lwd=3,xlab=expression(mu))
abline(v=5,lty=2,lwd=2,col=2)

sig2.sample = 1/phi[seq(1000,M,by=9)]

plot(density(sig2.sample,adjust=2),main="",
      lwd=3,xlab=expression(sigma^2))
abline(v=2,lty=2,lwd=2,col=2)

## Monte Carlo

mean(mu.sample)
# 5.157425

mean(sig2.sample)
# 2.619667

quantile(mu.sample, c(0.025,0.975))
#4.427828 5.819331

quantile(sig2.sample, c(0.025,0.975))
#1.404951 4.828969
```





Algoritmo de Metropolis-Hastings

- O amostrador de Gibbs é muito útil quando conhecemos a forma fechada das condicionais completas.

- Em muitas situações, essas distribuições não são conhecidas.

- O algoritmo de Metropolis generaliza o amostrador de Gibbs.

- Proposto originalmente por Metropolis et al. (1953) e generalizado por Hastings (1970), resultando no algoritmo de Metropolis-Hastings.

- É usado para gerar amostras de uma distribuição multivariada quando as condicionais completas são impossíveis de se amostrar diretamente.

- Ideia básica: gerar valores de uma distribuição proposta e aceitá-los ou rejeitá-los de acordo com uma regra de aceitação.

- Suponha que:
  - Estamos interessados em gerar amostras da distribuição alvo $p(\theta)$.
  - As distribuições condicionais completas são impossíveis de se amostrar diretamente.
  - Para cada parâmetro θ_j existe uma distribuição proposta $q(\theta_j \mid \cdot)$ que podemos gerar amostras facilmente e que podem ser usadas como alternativa à condicional completa.
- Dadas essas suposições, amostras de θ são obtidas usando o seguinte algoritmo:

1. Inicialize $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$
2. Para k de 1 até M, faça
   (a) Para j de 1 até p, faça
      i. Gere um valor proposto θ^*, usando
      $$\theta_j^{(k)} \sim q_j(\,.\,|\,\theta_j^{(k-1)})$$
      ii. Aceite o valor proposto com probabilidade
      $$\alpha = min\left\{1, \frac{p(\theta^* \,|\, y, \theta_{-j}^{(a)})\; q(\theta_j^{(k-1)} \,|\, \theta^*)}{p(\theta_j^{(k-1)} \,|\, y, \theta_{-j}^{(a)})\; q(\theta^* \,|\, \theta_j^{(k-1)})}\right\}$$
      onde $\theta_{-j}^{(a)} = (\theta_1^{(k)}, \ldots, \theta_{j-1}^{(k)}, \theta_{j+1}^{(k-1)}, \ldots, \theta_p^{(k-1)})$
3. Avalie a convergência

Resumo e Diagnóstico

- Para gerar amostras de uma distribuição usando as condicionais completas, podemos usar:
  - Amostrador de Gibbs: se soubermos gerar todas as condicionais completas.
  - Gibbs dentro do Metropolis: se soubermos gerar direto algumas condicionais completas, mas não outras.
  - Metropolis-Hastings: se não soubermos gerar diretamente das condicionais completas.
- Métodos via MCMC são muito úteis para resolver problemas práticos.
- É preciso garantir que as cadeias convergiram:
  - Traço das cadeias: gráfico de cada parâmetro ao longo das iterações.
  - Testes de convergência: Geweke, Raftery-Lewis, Gelman-Rubin.
  - Implementados no pacote `coda` do `R`.

Modelo de Regressão Linear Bayesiano

- Considere o modelo de regressão linear simples:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$$

onde β_0 é o intercepto, β_1 é o coeficiente angular e ϵ_i é o termo de erro aleatório.

- Suponha que os erros são independentes e normalmente distribuídos:

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n$$

- A função de verossimilhança é:

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2; \mathbf{y}) &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2 \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\} \end{aligned}$$

Modelo de Regressão Linear Bayesiano

- Suponha independência a priori: $p(\beta_0, \beta_1, \sigma^2) = p(\beta_0)p(\beta_1)p(\sigma^2)$, com

$$\beta_0 \sim N(\mu_{00}, \tau_0^2), \quad \beta_1 \sim N(\mu_{10}, \tau_1^2), \quad \sigma^2 \sim \text{Gama-Inv}(a, b)$$

- A posteriori conjunta é dada por:

$$\begin{aligned} p(\beta_0, \beta_1, \sigma^2 \mid \mathbf{y}) \propto (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\} \\ \times \exp\left\{ -\frac{1}{2\tau_0^2} (\beta_0 - \mu_{00})^2 \right\} \\ \times \exp\left\{ -\frac{1}{2\tau_1^2} (\beta_1 - \mu_{10})^2 \right\} \\ \times (\sigma^2)^{-a-1} \exp\left\{ -\frac{b}{\sigma^2} \right\} \end{aligned}$$

Modelo de Regressão Linear Bayesiano

Próximas etapas:

1. Derivar as Condicionais Completas a partir da posteriori conjunta.
2. Verificar Conhecimento das Condicionais: Se todas forem conhecidas, podemos prosseguir com Gibbs diretamente (é o caso do modelo de regressão linear normal).
3. Aplicar o Algoritmo de Gibbs para gerar amostras da posteriori conjunta.
4. Avaliar Convergência e Diagnósticos das Cadeias: Gráficos de traço e Testes de convergência (Geweke, Gelman-Rubin, Raftery-Lewis)
5. Realizar Inferência a partir das amostras geradas: estimativas de parâmetros, intervalos de credibilidade, previsões.