Introdução à Inferência Bayesiana

Démerson André Polli

ENAP - 03 a 12/12/2019 (aulas 01, 02 e 03)

Curso: Econometria (Séries Temporais Avançadas)

Professor: Démerson André Polli

O que é estatística (inferência)?

- Em diversas áreas de conhecimento humano a obtenção e descrição de dados é uma necessidade.
- Diversas situações impedem a obtenção de dados sobre toda a *população*.
  - *O custo de se observar toda a população pode ser proibitivo (ex. PNAD versus Censo Demográfico);*
  - *Alguns elementos da população podem não estar disponíveis no momento da coleta de dados (ex. alguns portadores de certa enfermidade são desconhecidos);*
  - *A observação de elementos amostrais causam a sua destruição (ex. uma fábrica de lâmpada precisa 'queimar' as lâmpadas para saber o tempo médio de durabilidade);*
  - *Alguns elementos da população somente estarão definidos no futuro (ex. valores não realizados de uma série temporal).*
- É necessário algum mecanismo para permitir *extrapolar* as informações observadas (na amostra) para a população.

Inferência estatística (I)

- A **população teórica** ou **população** é o universo dos elementos que se deseja caracterizar após a análise dos dados.
- Muitas vezes, é possível selecionar elementos apenas de parte da população. Este conjunto é chamado de **população acessível**.
- Um subconjunto de elementos da população (acessível) é selecionado. Tal conjunto é chamado **amostra**. Cada elemento deste conjunto é chamado de **elemento amostral**.
- A coleção de características observadas na amostra é chamada de **dados**. Cada característica observada nos elementos amostrais são chamados de **variável aleatória**.
- O comportamento das variáveis aleatórias em uma população é chamada de **distribuição**.

Inferência estatística (II)

Desta forma, observa-se dados em uma *amostra* com o objetivo de descrever o comportamento de *variáveis aleatórias* na *população*.

Como é feita, no entanto, a *ponte* entre a *amostra* e a *população*?

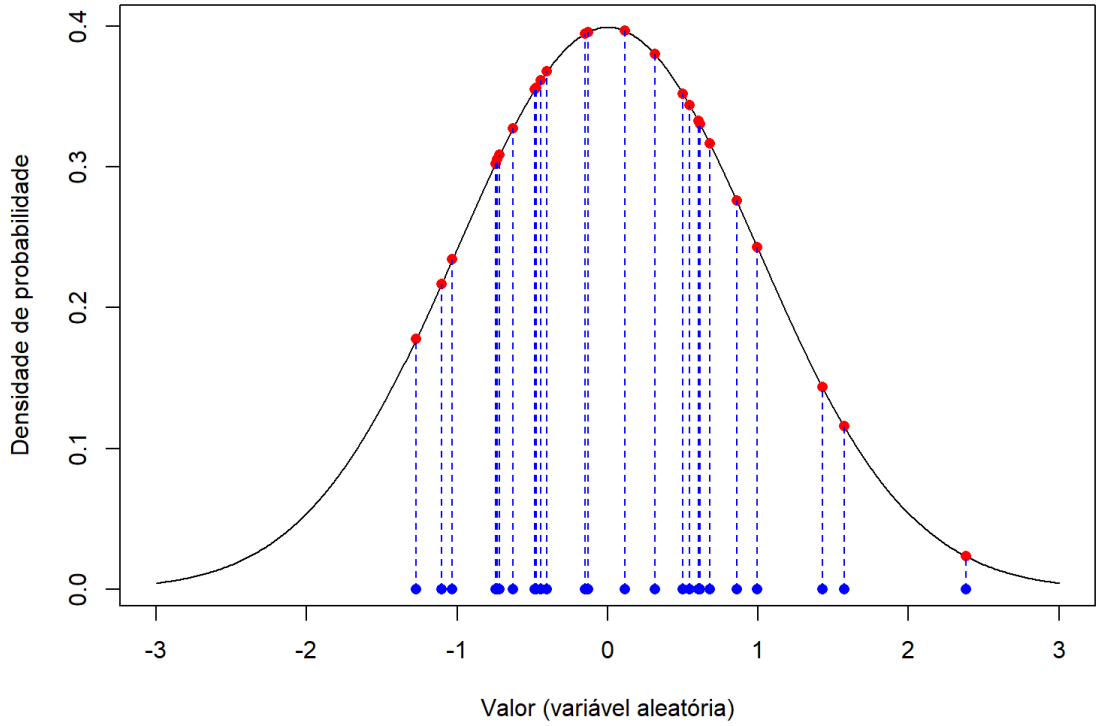
- Os valores observados (das variáveis aleatórias) nos elementos amostrais são descritos com relação aos valores observáveis, a frequência na qual ocorrem, etc.
- A *distribuição empírica* (histograma/tabelas de frequências) destas variáveis são aproximados de funções que descrevam:
  - *A* ***probabilidade*** *com a qual cada possível valor (da variável aleatória) ocorre.*
  - *A probabilidade está diretamente relacionada com a frequência observada na amostra.*

Como obter a *probabilidade* de um valor ocorrer (na população) se sabemos apenas a *frequência* com a qual tal valor ocorre na amostra?

Afinal, o que é *probabilidade*?

Inferência estatística (III)

A “seleção” de valores das variáveis aleatórias ocorrem da seguinte forma: cada realização amostral (ponto azul) ocorre de acordo com as respectivas probabilidades (pontos vermelhos) de ocorrência; os “valores” (pontos) na região de maior probabilidade ocorrem com maior “frequência” na amostra.



Inferência estatística (IV)

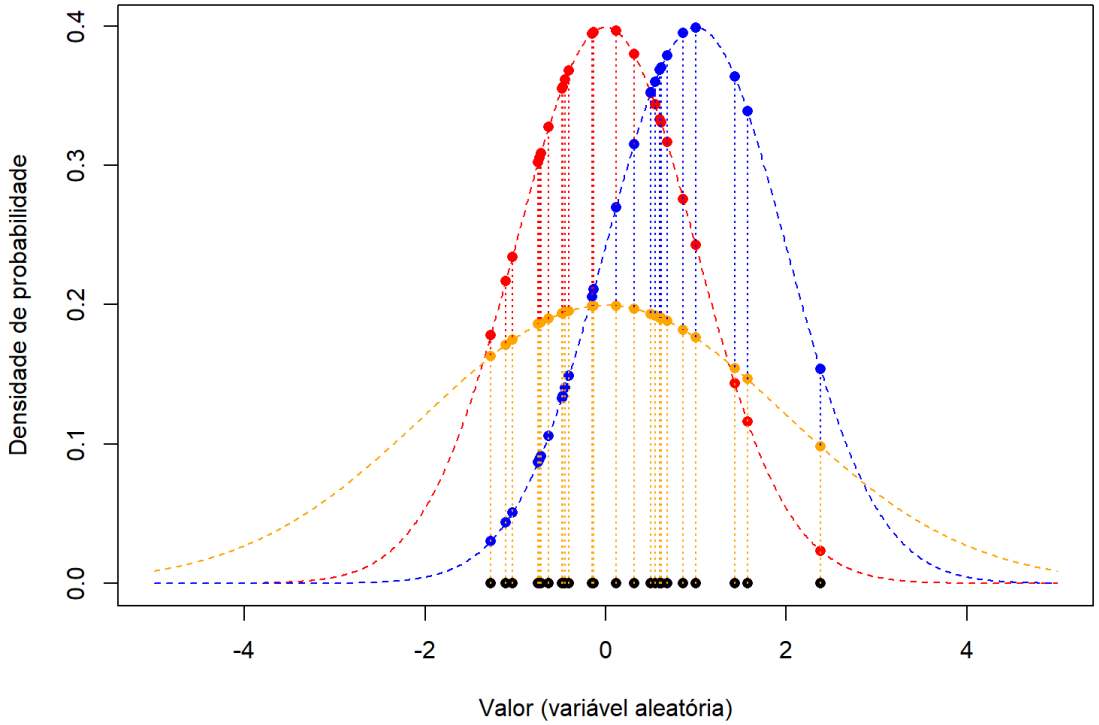
O gráfico do slide anterior demonstra a relação entre (i) os valores observados na amostra para as variáveis aleatórias e (ii) a forma funcional (matemática) que descreve a distribuição de probabilidade associada à estes valores.

Todo o trabalho estatístico consiste em:

- Observar os elementos amostrais (pontos azuis) através de algum método de seleção amostral (ou observação de um processo aleatório);
- Escolher a *forma funcional* que descreve a distribuição de probabilidades dos valores amostrais.
- Definir os *parâmetros* que modulam tal distribuição de probabilidade.

Inferência estatística (V)

Uma vez conhecida a *forma funcional* da distribuição de probabilidade é necessário definir os **parâmetros** que modulam tais formas funcionais. A combinação de forma funcional e parâmetros é o que define a **distribuição de probabilidade** relacionada com os dados. Diferentes distribuições podem ser plausíveis como as "geradoras" de uma amostra! Como escolher a distribuição mais **verossímil**?



Inferência clássica (I)

A **inferência clássica** se fundamenta no trabalho de Karl Pearson (1857 - 1936) e Ronald Fisher (1890 - 1962), bem como no conceito **frequentista** de probabilidade.

- A **probabilidade** é a proporção limite com a qual um evento ocorre (ou um valor é observado). Se um experimento aleatório for repetido "infinitas vezes", a proporção com que o evento de interesse (ou valor) é observado é a probabilidade de ocorrência deste evento (valor).

- A **distribuição de probabilidade** de uma variável aleatória possui forma funcional indexada por um parâmetro. Tal parâmetro é um valor desconhecido (*para* - além de, *metros* - medida) e **fixo**.

- Na *inferência classica* os *parâmetros* são *fixos* e, desta forma, não são aleatórios. Isto implica em algo muito importante: **os parâmetros são constantes a serem estimadas**.

Inferência clássica (II)

Na inferência clássica, os dados $X_1, X_2, \ldots, X_n$ possuem uma distribuição de probabilidade descrita pela forma funcional $f(X_i|\boldsymbol{\theta})$, $i = 1, 2, \ldots, n$, indexada por um vetor de parâmetros $\boldsymbol{\theta}$ fixo e desconhecido.

Uma vez observada uma amostra $x_1, x_2, \ldots, x_n$, supondo a **independência** e a **igualdade de distribuição**n entre as observações, define-se a função

$$g(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}).$$

Considerando que o vetor de parâmetros $\boldsymbol{\theta}$ é fixo mas *conhecido* e que $g(\cdot)$ é uma função do vetor de observações (amostra), $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$, então a **probabilidade conjunta** é definida por

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|\boldsymbol{\theta}) = g(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}).$$

Inferência clássica (III)

Considerando que o vetor de parâmetros $\boldsymbol{\theta}$ é fixo e *desconhecido*, fixada uma amostra $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$, e que $g(\cdot)$ é uma função do vetor de parâmetros, $\boldsymbol{\theta}$, então a **função de verossimilhança** é definida por

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = g(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}).$$

A **estimativa** do vetor de parâmetros pode ser obtida por técnicas como **mínimos quadrados** ou pela maximização da função de verossimilhanças

$$\widehat{\boldsymbol{\theta}} = \arg\max L(\boldsymbol{\theta}|\boldsymbol{x}).$$

Esta última técnica é conhecida por **método da máxima verossimilhança**.

Probabilidade condicional (I)

- A *probabilidade* é uma medida (função) que assume valores entre 0 e 1.
- O universo dos valores possíveis de uma variável aleatória é chamado de **espaço amostral** e é representado pela letra grega ômega maiúscula (Ω).
- A probabilidade de se observar um valor do espaço amostral é 1, ou seja, $\mathbb{P}(\Omega) = 1$.
- A probabilidade de um evento E representa "o tamanho" deste evento em relação ao espaço amostral:

$$\mathbb{P}(E) = \frac{\mathbb{P}(E)}{\mathbb{P}(\Omega)} = \frac{\mathbb{P}(E)}{1}.$$

Qual a probabilidade de ocorrer um evento A considerando que um evento (relacionado) B ocorreu?

Probabilidade condicional (II)

- Se um evento B relacionado com o evento de interesse A ocorreu:
  - *O evento A somente poderá ocorrer se ocorrer a intersecção entre A e B ($A \cap B$).*
  - *O universo de resultados possíveis não é mais o espaço amostral (Ω), mas o conjunto que representa o evento B.*
- A probabilidade do evento A ocorrer após a ocorrência do evento B é o "tamanho" da intersecção de A e B em relação ao "tamanho" do evento B:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- A expressão acima é chamada de **probabilidade condicional**: probabilidade de A condicional à (ocorrência de) B.

Fórmula de Bayes (I)

Observe que

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \rightarrow \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(A \cap B).$$

De forma análoga, tomando a probabilidade de B condicional à ocorrência de A,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \rightarrow \mathbb{P}(B|A) \cdot \mathbb{P}(A) = \mathbb{P}(A \cap B).$$

Fórmula de Bayes (II)

Igualando as duas expressões,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A),$$

e observando que

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)$$

(a probabilidade de B é a probabilidade de B ocorrer "junto" com A $[B \cap A]$, somada com a probabilidade de B ocorrer "separado" de A - $[B \cap A^c]$), obtém-se

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)}.$$

Tal expressão é chamada **fórmula de Bayes**.

Fórmula de Bayes (III)

Vejamos um exemplo. Suponha que o evento A seja "o(a) aluno(a) estudou para a prova" e que o evento B seja o evento "o(a) aluno(a) obteve nota maior que 5,0". Qual a probabilidade do(a) aluno(a) ter estudado para a prova considerando que sua nota foi maior que 5,0?

Suponha que a probabilidade de obter nota superior a 5,0 quando a pessoa estudou seja $\mathbb{P}(B|A) = 0.80$, a probabilidade de obter nota superior a 5,0 quando a pessoa não estudou seja $\mathbb{P}(B|A^c) = 0.10$ e que $1/4$ dos alunos estudam para as provas. Assim, lembrando que

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)},$$

temos

$$\mathbb{P}(A|B) = \frac{0.80 \cdot 0.25}{0.80 \cdot 0.25 + 0.10 \cdot 0.75} = \frac{0.200}{0.200 + 0.075} = \frac{0.200}{0.275}.$$

Logo, a probabilidade do(a) aluno(a) ter estudado considerando que sua nota foi maior que 5,0 é $\mathbb{P}(A|B) = 0.7273$.

Inferência bayesiana (I)

A **inferência bayesiana** se fundamenta no trabalho de Thomas Bayes (1701 - 1761), Harold Jeffreys (1891 - 1989) e Bruno de Finetti (1906 - 1985). Na inferência bayesiana a probabilidade é uma **medida de incerteza**.

- A **probabilidade** é uma medida de incerteza a respeito da ocorrência de um evento (ou observação de um valor). A probabilidade pode ser definida de forma subjetiva (o quanto você acredita que tal evento ocorrerá?), mas é atualizável pelos dados.

- Tal como na inferência clássica, a **distribuição de probabilidade** de uma variável aleatória possui forma funcional indexada por um parâmetro desconhecido. Como o parâmetro é desconhecido, **atribui-se ao parâmetro uma distribuição de probabilidade**.

- **Na inferência bayesiana os parâmetros são aleatórios**.

Inferência bayesiana (II)

- A distribuição de probabilidade dos parâmetros é atribuída antes de se observar a amostra. Tal distribuição é chamada **distribuição a priori**.
- Toda informação contida na amostra é representada pela **função de verossimilhança**.
- A distribuição de probabilidade dos parâmetros é atualizada após observar os elementos amostrais. A distribuição de probabilidade dos parâmetros após a observação da amostra é chamada **distribuição a posteriori**.
- Toda **inferência** estatística (sob a teoria bayesiana) é realizada (i) obtendo-se valores de parâmetros da *distribuição a posteriori* e, se necessário, (ii) calculando-se a probabilidade dos valores amostrais considerando que o parâmetro está fixado.

Inferência bayesiana (III)

Na inferência bayesiana, os dados $X_1, X_2, \ldots, X_n$ possuem uma distribuição de probabilidade descrita pela forma funcional $f(X_i|\boldsymbol{\theta})$, $i = 1, 2, \ldots, n$, indexada por um vetor de parâmetros $\boldsymbol{\theta}$ aleatório e desconhecido (não observável).

Uma vez observada uma amostra $x_1, x_2, \ldots, x_n$, define-se a **função de verossimilhança**,

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}).$$

Aplicando a *fórmula de Bayes* para obter a distribuição de $\boldsymbol{\theta}$ condicional aos dados ($\boldsymbol{X}$), e lembrando a equivalencia da distribuição de probabilidade conjunta dos valores amostrais com a função de verossimilhança, segue que

$$\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X} = \boldsymbol{x}) = \frac{L(\boldsymbol{\theta}|\boldsymbol{x}) \times \mathbb{P}(\boldsymbol{\theta})}{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x})}.$$

Inferência bayesiana (IV)

A distribuição $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X} = \boldsymbol{x})$ é a **distribuição a posteriori** de $\boldsymbol{\theta}$ dado $\boldsymbol{X}$, representada por $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, $\mathbb{P}(\boldsymbol{\theta})$ é a **distribuição a priori** de $\boldsymbol{\theta}$, representada por $\pi(\boldsymbol{\theta})$, e $\mathbb{P}(\boldsymbol{X} = \boldsymbol{x})$ é a **distribuição preditiva** de $\boldsymbol{X}$, representada por $f(\boldsymbol{x})$. Assim,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{\theta}|\boldsymbol{x}) \times \pi(\boldsymbol{\theta})}{f(\boldsymbol{x})} \propto L(\boldsymbol{\theta}|\boldsymbol{x}) \times \pi(\boldsymbol{\theta}).$$

A distribuição *a posteriori* é proporcional ao produto da *verossimilhança* com a distribuição *a priori*. A *verossimilhança* carrega as informações da amostra e a *distribuição a priori* carrega a informação prévia (antes dos dados). Assim, a *distribuição a posteriori* pondera as duas fontes de informação.

Inferência bayesiana (V)

Um apostador ganha R$ 1,00 cada vez que lança uma moeda e obtém *cara*. Qual a probabilidade de ganhar o prêmio:

- Antes de jogar alguma vez?
- Após observar o resultado de 1, 2 e 3 lançamentos?

A contagem de *caras* em n lançamentos segue uma distribuição binomial, com **função de verossimilhança**

$$L(\theta|\boldsymbol{x}) \propto \theta^x (1-\theta)^{n-x}.$$

O parâmetro θ pode ser descrito (a priori) com uma distribuição Dirichlet, dada por

$$\pi(\theta) \propto \theta^{a-1} (1-\theta)^{b-1}$$

A distribuição *a posteriori* é dada por

$$\pi(\theta|\boldsymbol{x}) \propto \theta^x (1-\theta)^{n-x} \cdot \theta^{a-1} (1-\theta)^{b-1} = \theta^{x+a-1} (1-\theta)^{n-x+b-1}.$$

Inferência bayesiana (VI)

Na distribuição a priori, $\pi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$, os valores $a \in \mathbb{R}$ e $b \in \mathbb{R}$ são chamados **hiper-parâmetros**.

Tomando os hiper-parâmetros $a = b = 1$ obtém-se a uma distribuição a priori uniforme (atribui probabilidade igual para qualquer possível valor de θ). Esta distribuição a priori representa a completa ignorância a respeito do parâmetro θ.

Se a priori é uniforme, a distribuição a posterior após observar *cara* será:

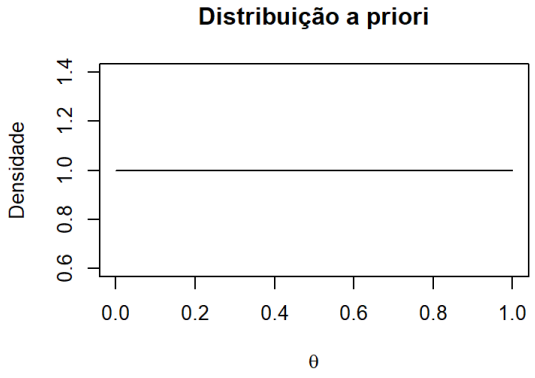
$$\pi(\theta|x_1 = 1) \propto \theta^0(1-\theta)^0 \cdot \theta^1(1-\theta)^0 = \theta^1(1-\theta)^0 = \theta.$$

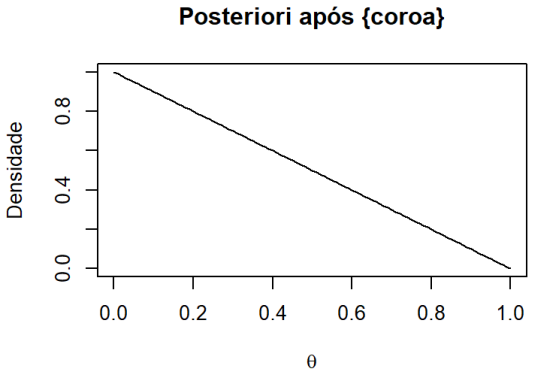
Se a priori é uniforme, a distribuição a posterior após observar *coroa* será:

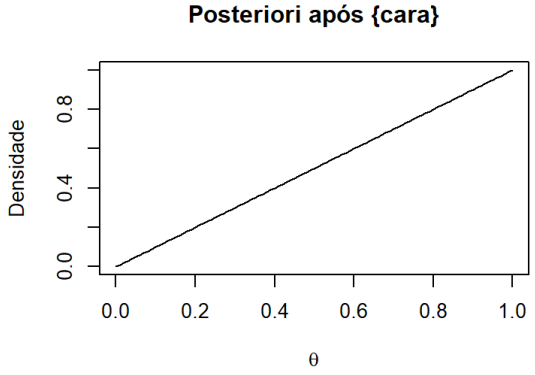
$$\pi(\theta|x_1 = 0) \propto \theta^0(1-\theta)^0 \cdot \theta^0(1-\theta)^1 = \theta^0(1-\theta)^1 = 1-\theta.$$

No próximo slide se apresentam os diferentes gráficos da priori e posterioris.

Inferência bayesiana (VII)

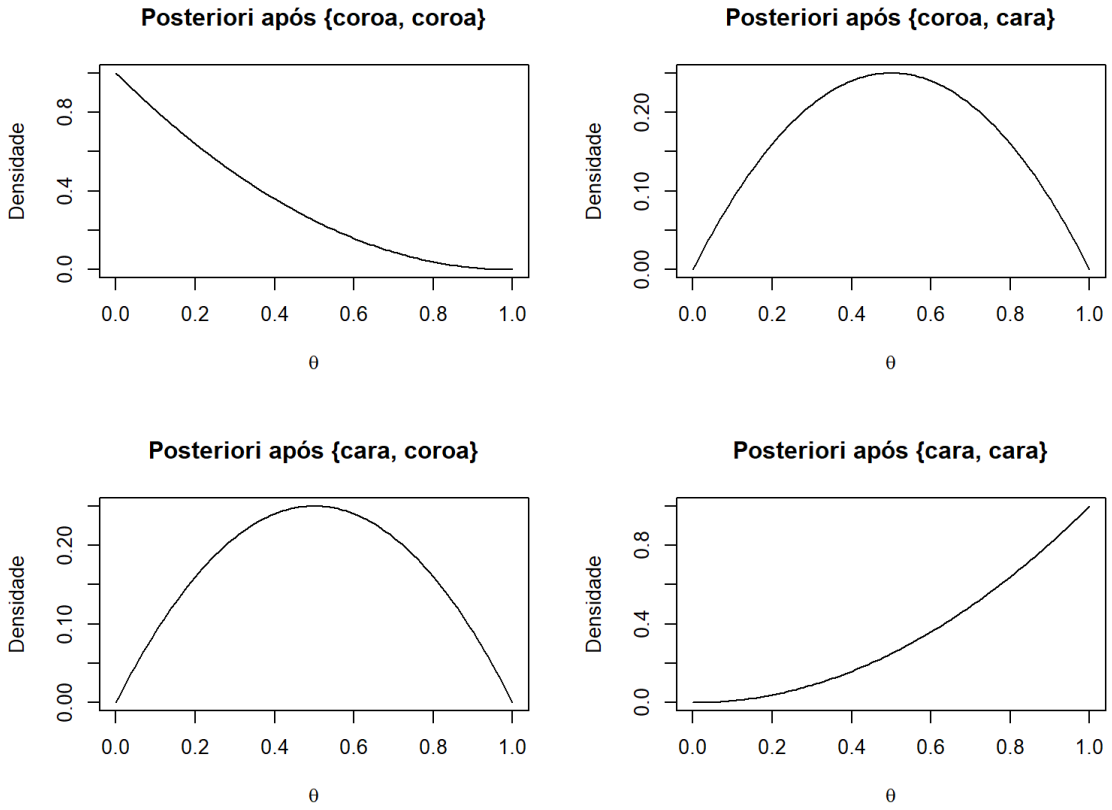






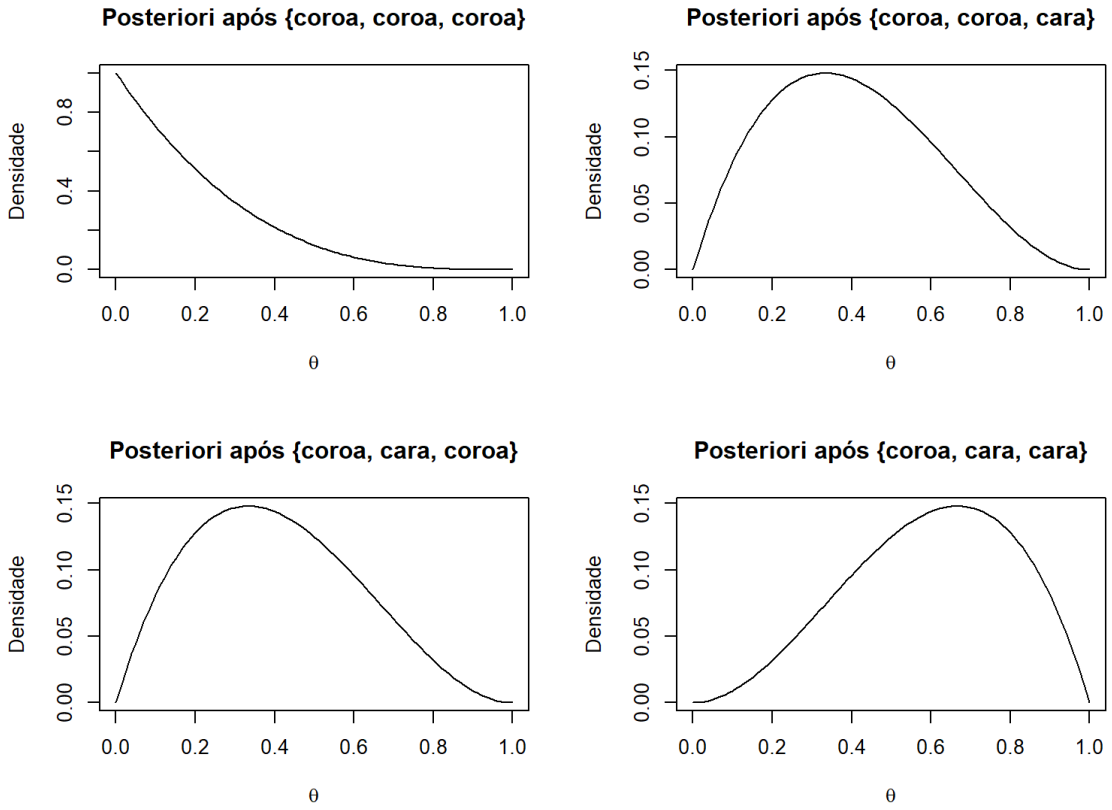
Inferência bayesiana (VIII)

Posteriori no segundo passo: $\pi(\theta|x_1, x_2) \propto \theta^{x_1+x_2}(1-\theta)^{2-x_1-x_2}$



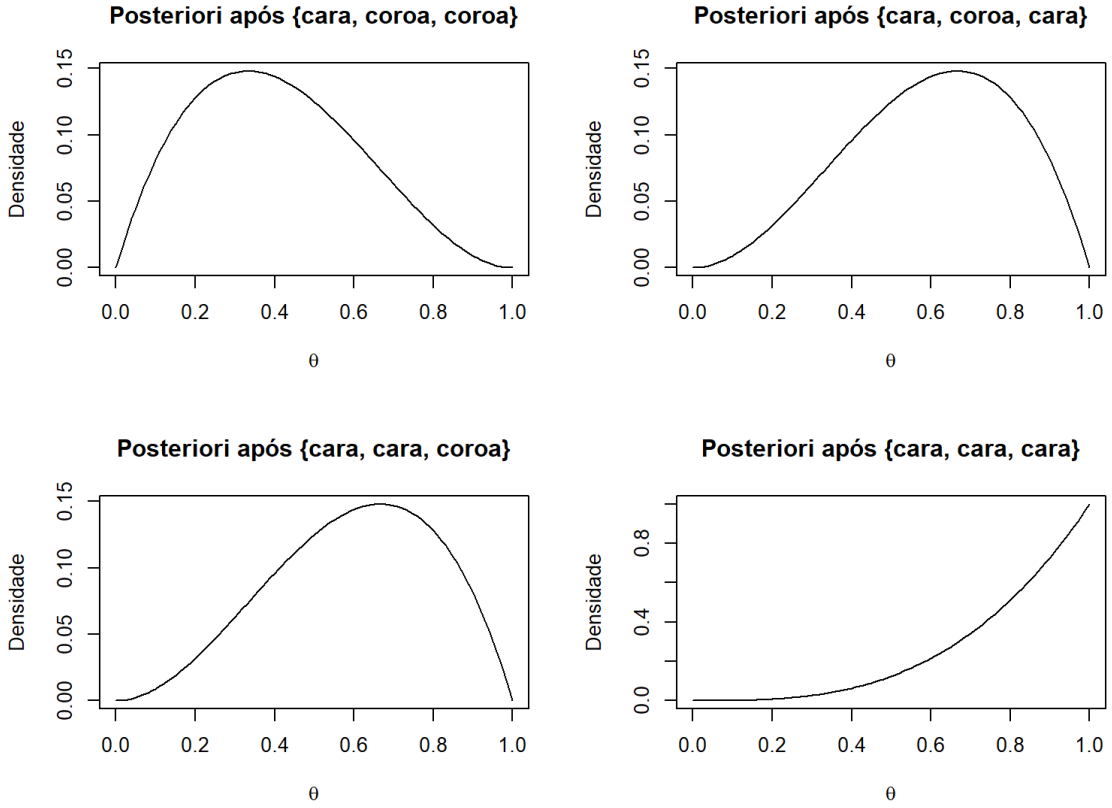
Inferência bayesiana (IX)

Posteriori no terceiro passo: $\pi(\theta|x_1, x_2, x_3) \propto \theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3}$.



Inferência bayesiana (X)

Posteriori no terceiro passo: $\pi(\theta|x_1, x_2, x_3) \propto \theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3}$.



Inferência bayesiana (XI)

- Toda informação da amostra está contida na verossimilhança e a *distribuição a posteriori* é proporcional à verossimilhança.
- **Teorema de 'de Finetti':** a inferência não deve ser afetada pela *ordem* em que os valores amostrais são observados (**permutabilidade**) - isto é observável nos gráficos das páginas anteriores.
- **Princípio da verossimilhança:** amostras com verossimilhanças proporcionais (ou iguais) devem resultar na mesma inferência.
- A estatística clássica não respeita o princípio da verossimilhança (ex. uma sequência {coroa, coroa, cara} resulta em inferências distintas se for considerado que X segue uma distribuição binomial ou uma distribuição geométrica).
- A *permutabilidade* de 'de Finetti' é uma condição menos restritiva de que a independência (exigida na maioria dos modelos da inferência clássica).

Priori imprópria

Uma priori $\pi(\boldsymbol{\theta})$ é dita ser uma **priori imprópria** se

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \pi(\boldsymbol{\theta})\mathrm{d}\,\boldsymbol{\theta} = \infty.$$

Em diversas situações, apesar da priori ser imprópria, a posteriori poderá ser própria. No entanto, nestas condições é necessário verificar.

O uso de prioris impróprias é interessante para se definir **prioris não informativas**. Um exemplo de tal priori será mostrado no próximo slide

Priori de Jeffreys

A **priori de Jeffreys** é usada para se definir prioris não informativas. Tal priori é proporcional à *raiz quadrada do determinante da matriz de informação de Fisher* associada ao *parâmetro* $\boldsymbol{\theta}$ na distribuição dos dados $\boldsymbol{X}|\boldsymbol{\theta}$.

É comum que a priori de Jeffreys seja uma **priori imprópria**. No entanto, a posteriori, na maioria das vezes, será própria.

Priori conjugada

Uma priori é dita ser **conjugada** quando a posteriori $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ é da mesma família de distribuição da priori $\pi(\boldsymbol{\theta})$.

No exemplo do lançamento da moeda, a priori de θ é uma *Dirichlet* cuja função de distribuição é dada por

$$\pi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}.$$

Naquele caso foi tomado $a = 1$ e $b = 1$ para definir uma "*priori não informativa*".

Após observar k "caras" em n lançamentos da moeda, a posteriori obtida é

$$\pi(\theta|x_1, x_2, \ldots, x_n) \propto \theta^{k+a-1}(1-\theta)^{n-k+b-1}.$$

Tal distribuição é *Dirichlet* com parâmetros $a^\star = k + a$ e $b^\star = n - k + b$.

Econometria bayesiana (I)

Um modelo de **regressão linear** é definido tal que

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \xi_i,\ \xi_i \sim N(0, \sigma^2),$$

ou seja,

$$Y \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \xi_i, \sigma^2),$$

É usual atribuir ao vetor $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]$ uma priori *Normal multivariada* e ao parâmetro σ uma distribuição de Jeffreys (imprópria) $\pi(\sigma) \propto 1/\sigma$.

Em um modelo de regressão, suponha que $\boldsymbol{\beta}$ seja distribuído de acordo com uma *Normal* p*-variada* de média $\boldsymbol{\eta}_0$ e variância $\sigma^2 \boldsymbol{\Sigma}_0$, e que σ^2 seja distribuído de acordo com uma *Gama invertida* com parâmetros $\alpha_0/2$ e $\delta_0/2$.

Neste caso, a posteriori $\sigma^2 | \boldsymbol{Y}$ segue uma distribuição *Gama invertida* com parâmetros $\alpha_1/2$ e $\delta_1/2$ dados por

$$\alpha_1 = \alpha_0 + n,\ \delta_1 = \delta_0 + \boldsymbol{y}'\boldsymbol{y} + \boldsymbol{\eta}_0' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\eta}_0 - \boldsymbol{\eta}_1' \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\eta}_1$$

Econometria bayesiana (II)

Na regressão linear, a posteriori $\boldsymbol{\beta}|\boldsymbol{Y}, \sigma^2$ segue uma distribuição *t-Student* p*-variada* com parâmetros $(\alpha_1, \boldsymbol{\eta}_1, (\delta_1/\alpha_1)\boldsymbol{\Sigma}_1)$ em que

$$\boldsymbol{\eta}_1 = \boldsymbol{\Sigma}_1(\boldsymbol{x}'\boldsymbol{y} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)$$

e

$$\boldsymbol{\Sigma}_1 = (\boldsymbol{x}'\boldsymbol{x} + \boldsymbol{\Sigma}^{-1})^{-1}$$

Econometria bayesiana (III)

Exemplo de regressão linear (clássica):

```
## Annette Dobson (1990) "An Introduction to Generalized Linear Models".
## Page 9: Plant Weight Data.
ctl <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
trt <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
group <- gl(2, 10, 20, labels = c("Ctl","Trt"))
weight <- c(ctl, trt)
lm.D9 <- lm(weight ~ group)
lm.D90 <- lm(weight ~ group - 1) # omitting intercept
```

Econometria bayesiana (IV)

```r
summary(lm.D9)
```

```
##
## Call:
## lm(formula = weight ~ group)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0710 -0.4938  0.0685  0.2462  1.3690
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.0320     0.2202  22.850 9.55e-15 ***
## groupTrt     -0.3710     0.3114  -1.191    0.249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6964 on 18 degrees of freedom
## Multiple R-squared:  0.07308,    Adjusted R-squared:  0.02158
## F-statistic: 1.419 on 1 and 18 DF,  p-value: 0.249
```

Econometria bayesiana (V)

```
summary(lm.D90)
```

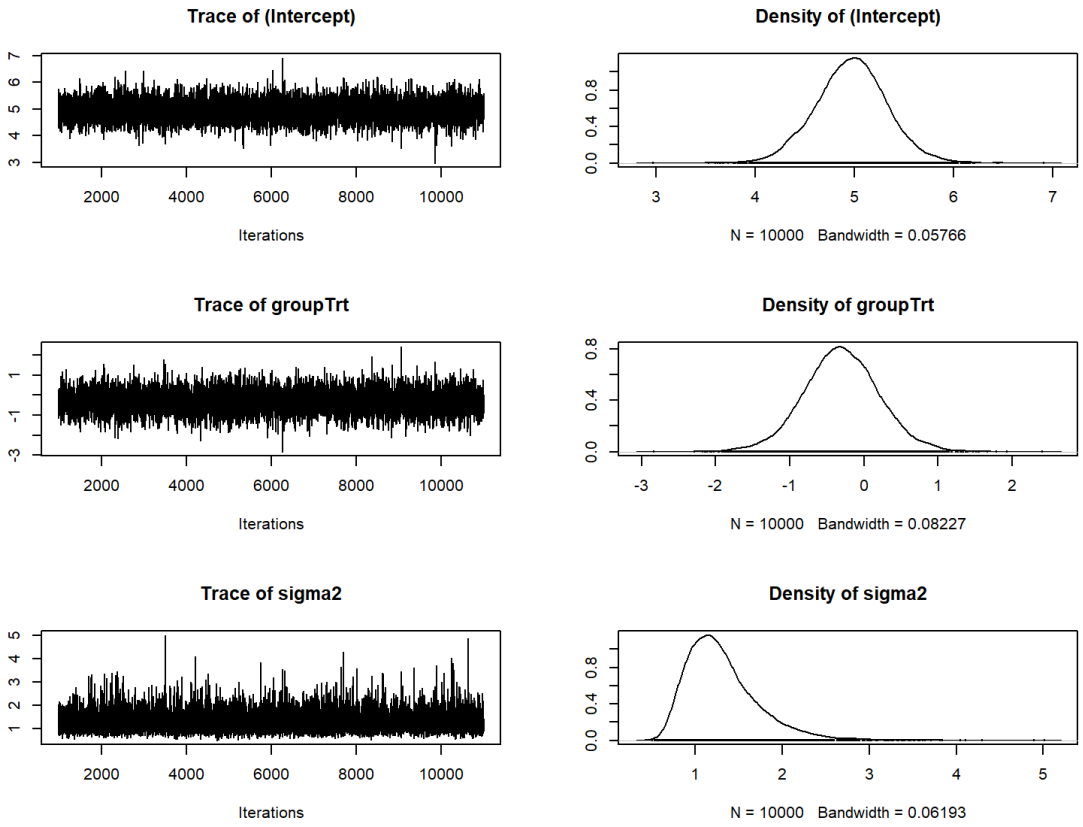
```
##
## Call:
## lm(formula = weight ~ group - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0710 -0.4938  0.0685  0.2462  1.3690
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## groupCtl   5.0320     0.2202   22.85 9.55e-15 ***
## groupTrt   4.6610     0.2202   21.16 3.62e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6964 on 18 degrees of freedom
## Multiple R-squared:  0.9818, Adjusted R-squared:  0.9798
## F-statistic: 485.1 on 2 and 18 DF,  p-value: < 2.2e-16
```

Econometria bayesiana (VI)

```
library(MCMCpack)
posterior = MCMCregress(weight ~ group,
                        b0 = 0,
                        B0 = 0.1,
                        sigma.mu = 5,
                        sigma.var = 25)
```

Econometria bayesiana (VII)

```
plot(posterior)
```



Econometria bayesiana (VIII)

- The number of iterations required to estimate the quantile q to within an accuracy of $\pm r$ with probability p is calculated. Separate calculations are performed for each variable within each chain.
- The minimum length is the required sample size for a chain with no correlation between consecutive samples. Positive autocorrelation will increase the required sample size above this minimum value.
- An estimate I (the 'dependence factor') of the extent to which autocorrelation inflates the required sample size is also provided.
- Values of I larger than 5 indicate strong autocorrelation which may be due to a poor choice of starting value, high posterior correlations or 'stickiness' of the MCMC algorithm.
- The number of 'burn in' iterations to be discarded at the beginning of the chain is also calculated.

Econometria bayesiana (IX)

```
raftery.diag(posterior, q = 0.025, r = 0.005, s = 0.95, converge.eps = 0.001)
```

```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
##             Burn-in  Total Lower bound  Dependence
##             (M)      (N)   (Nmin)       factor (I)
##  (Intercept) 2        3802  3746         1.010
##  groupTrt    2        3680  3746         0.982
##  sigma2      2        3710  3746         0.990
```

Econometria bayesiana (X)

```r
summary(posterior)
```

```
##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                Mean     SD Naive SE Time-series SE
## (Intercept)  4.9675 0.3587 0.003587       0.003511
## groupTrt    -0.3035 0.5070 0.005070       0.005070
## sigma2       1.3102 0.4198 0.004198       0.004506
##
## 2. Quantiles for each variable:
##
##                2.5%     25%     50%     75%  97.5%
## (Intercept)  4.2599  4.7399  4.9723 5.19973 5.6653
## groupTrt    -1.3186 -0.6317 -0.3079 0.02452 0.7101
## sigma2       0.7343  1.0173  1.2340 1.51126 2.3193
```

Econometria bayesiana (XI)

O conjunto *birthwt* apresenta fatores de risco associados com o risco de um bebê nascer com menos que 2,5 Kg. O exercício é estimar a probabilidade de baixo peso da criança estimada pela idade e etnia da mãe e da indicadora de que a mãe fumou durante a gravidez.

```
data(birthwt)
birthwt$race = factor(birthwt$race, labels = c("white", "black", "other"))
head(birthwt)
```

```
##    low age lwt  race smoke ptl ht ui ftv  bwt
## 85   0  19 182 black     0   0  0  1   0 2523
## 86   0  33 155 other     0   0  0  0   3 2551
## 87   0  20 105 white     1   0  0  0   1 2557
## 88   0  21 108 white     1   0  0  1   2 2594
## 89   0  18 107 white     1   0  0  1   0 2600
## 91   0  21 124 other     0   0  0  0   0 2622
```

Econometria bayesiana (XII)

Na primeira regressão a *distribuição a priori* dos coeficientes de regressão é **Normal multivariada**. Assume-se *a priori* que os coeficientes distribuídos de acordo com uma Normal de média $b_0 = 0.0$ e variância $\sigma_0^2 = 1/0.001 = 1000$.

Observe que a variância na priori é bastante grande. Isto é uma forma de informar pouco conhecimento a priori sobre o parâmetro (é uma distribuição pouco informativa). A priori é uma distribuição centrada em 0 mas que atribui alta probabilidade para uma faixa bastante extensa de valores.

```
posterior <- MCMClogit(low ~ age + race + smoke, b0 = 0, B0 = .001,
                       data = birthwt, burnin = 5000, mcmc = 30000)
```

Econometria bayesiana (XIII)

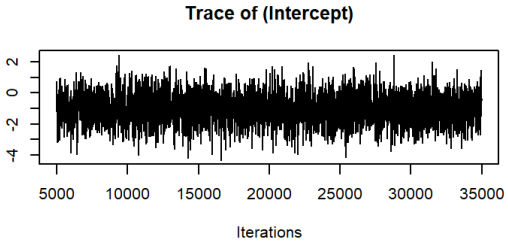
Resultados para o modelo com *priori Normal multivariada*:

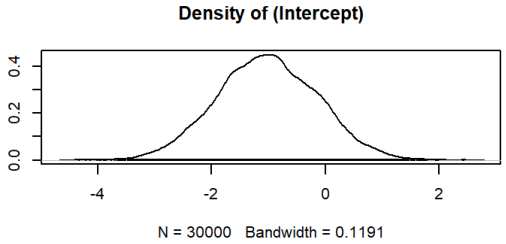
```
raftery.diag(posterior, q = 0.025, r = 0.005, s = 0.95, converge.eps = 0.001)
```

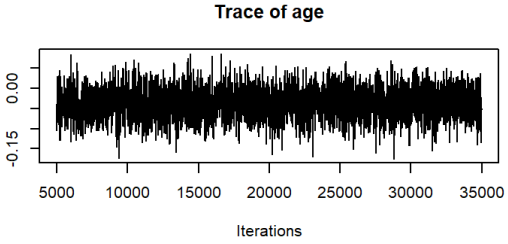
```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
##              Burn-in  Total Lower bound  Dependence
##              (M)      (N)   (Nmin)       factor (I)
##  (Intercept) 27       29362 3746         7.84
##  age         33       37155 3746         9.92
##  raceblack   26       28129 3746         7.51
##  raceother   27       29128 3746         7.78
##  smoke       28       30214 3746         8.07
```

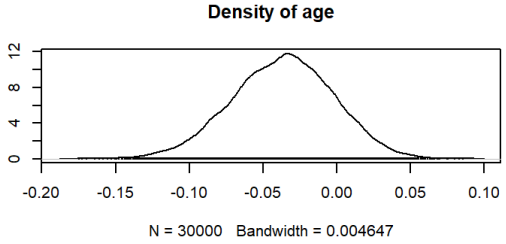
Econometria bayesiana (XIV)

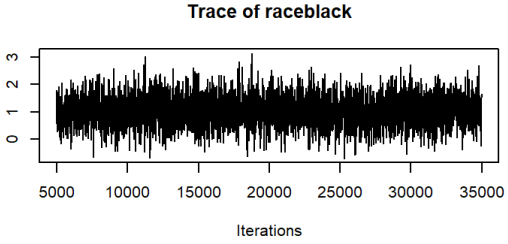
Resultados para o modelo com *priori Normal multivariada*:

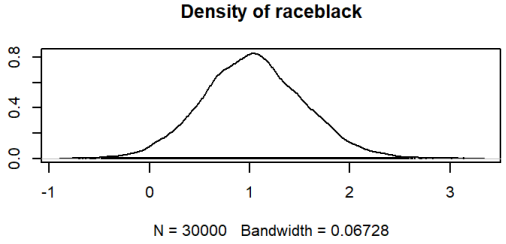

Trace of (Intercept)
-4 -2 0 2
5000 10000 15000 20000 25000 30000 35000
Iterations

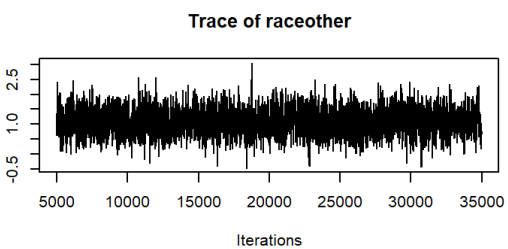

Density of (Intercept)
0.0 0.2 0.4
-4 -2 0 2
N = 30000 Bandwidth = 0.1191

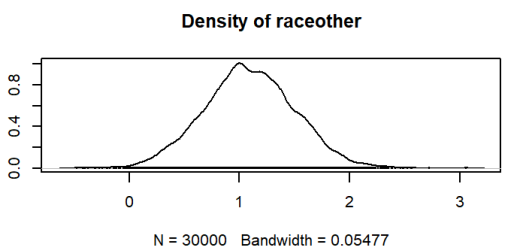

Trace of age
-0.15 0.00
5000 10000 15000 20000 25000 30000 35000
Iterations

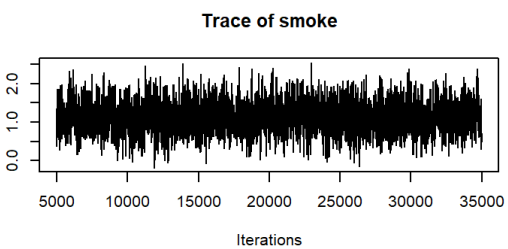

Density of age
0 4 8 12
-0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10
N = 30000 Bandwidth = 0.004647

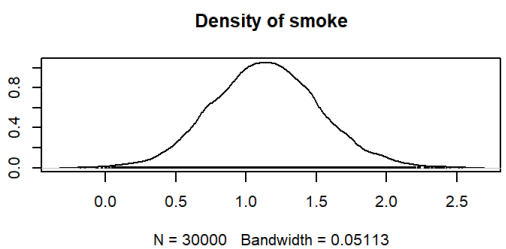

Trace of raceblack
0 1 2 3
5000 10000 15000 20000 25000 30000 35000
Iterations


Density of raceblack
0.0 0.4 0.8
-1 0 1 2 3
N = 30000 Bandwidth = 0.06728


Trace of raceother
-0.5 1.0 2.5
5000 10000 15000 20000 25000 30000 35000
Iterations


Density of raceother
0.0 0.4 0.8
0 1 2 3
N = 30000 Bandwidth = 0.05477


Trace of smoke
0.0 1.0 2.0
5000 10000 15000 20000 25000 30000 35000
Iterations


Density of smoke
0.0 0.4 0.8
0.0 0.5 1.0 1.5 2.0 2.5
N = 30000 Bandwidth = 0.05113

Econometria bayesiana (XV)

Resultados para o modelo com *priori Normal multivariada*:

```
##
## Iterations = 5001:35000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 30000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                 Mean      SD  Naive SE Time-series SE
## (Intercept) -1.02466 0.88285 0.0050971       0.020671
## age         -0.03677 0.03445 0.0001989       0.000818
## raceblack    1.03557 0.50278 0.0029028       0.011496
## raceother    1.08952 0.41094 0.0023726       0.009754
## smoke        1.14235 0.38083 0.0021987       0.008868
##
## 2. Quantiles for each variable:
##
##                2.5%      25%     50%      75%   97.5%
## (Intercept) -2.74012 -1.62571 -1.0300 -0.41900 0.71636
## age         -0.10599 -0.05975 -0.0356 -0.01338 0.02832
## raceblack    0.06208  0.69663  1.0265  1.36513 2.05357
## raceother    0.29086  0.81898  1.0813  1.36320 1.90013
## smoke        0.41147  0.88567  1.1405  1.39374 1.92257
```

Econometria bayesiana (XVI)

É possível, no entanto, fornecer uma priori diferente do padrão. No exemplo abaixo, cada coeficiente de regressão segue uma distribuição de Cauchy e são independentes.

```
logpriorfun <- function(beta){
  sum(dcauchy(beta, log=TRUE))
}

posterior <- MCMClogit(low ~ age + race + smoke,
                          data = birthwt, user.prior.density = logpriorfun,
                          logfun = TRUE, burnin = 5000, mcmc = 50000)
```

```
##
##
## @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
## The Metropolis acceptance rate was 0.23838
## @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
```

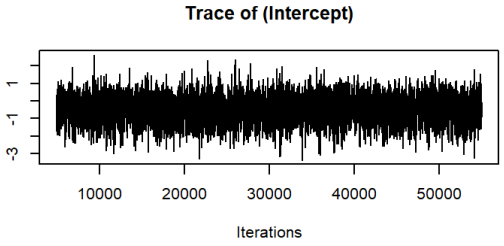
Econometria bayesiana (XVII)

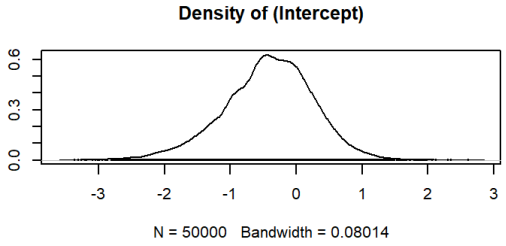
Resultados para o modelo com *priori Cauchy*:

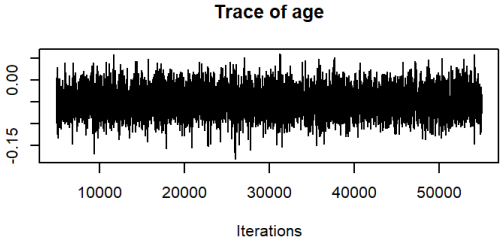
```
##
## Quantile (q) = 0.025
## Accuracy (r) = +/- 0.005
## Probability (s) = 0.95
##
##             Burn-in  Total Lower bound  Dependence
##             (M)      (N)   (Nmin)       factor (I)
##  (Intercept) 30       31334 3746         8.36
##  age         28       30222 3746         8.07
##  raceblack   28       29997 3746         8.01
##  raceother   28       30527 3746         8.15
##  smoke       29       31690 3746         8.46
```

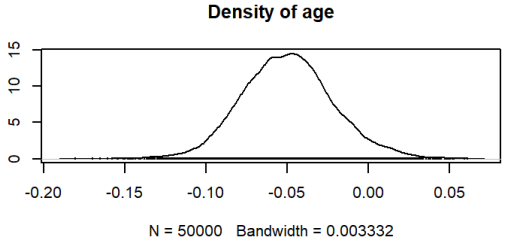
Econometria bayesiana (XVIII)

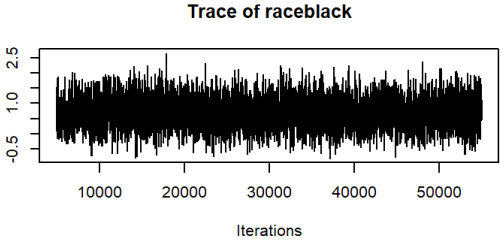
Resultados para o modelo com *priori Cauchy*:

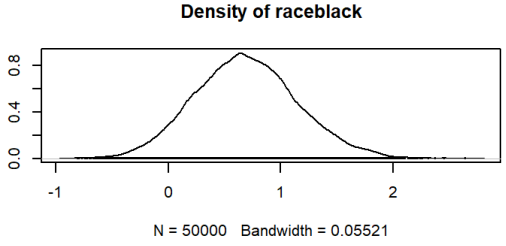

Trace of (Intercept)
-3
-1
1
10000
20000
30000
40000
50000
Iterations

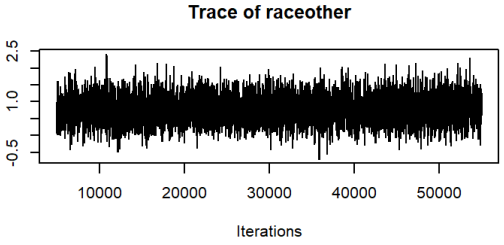

Density of (Intercept)
0.0
0.3
0.6
-3
-2
-1
0
1
2
3
N = 50000 Bandwidth = 0.08014

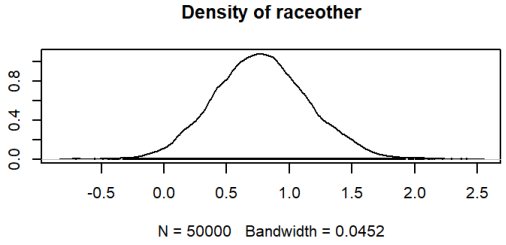

Trace of age
-0.15
0.00
10000
20000
30000
40000
50000
Iterations

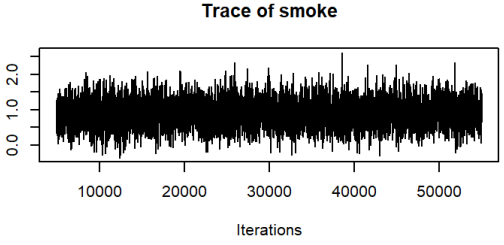

Density of age
0
5
10
15
-0.20
-0.15
-0.10
-0.05
0.00
0.05
N = 50000 Bandwidth = 0.003332

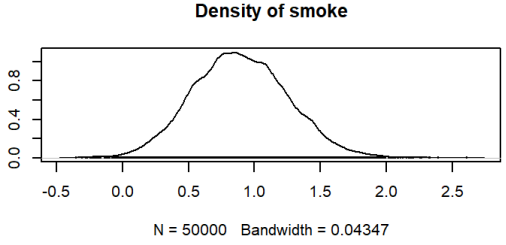

Trace of raceblack
-0.5
1.0
2.5
10000
20000
30000
40000
50000
Iterations


Density of raceblack
0.0
0.4
0.8
-1
0
1
2
N = 50000 Bandwidth = 0.05521


Trace of raceother
-0.5
1.0
2.5
10000
20000
30000
40000
50000
Iterations


Density of raceother
0.0
0.4
0.8
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
N = 50000 Bandwidth = 0.0452


Trace of smoke
0.0
1.0
2.0
10000
20000
30000
40000
50000
Iterations


Density of smoke
0.0
0.4
0.8
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
N = 50000 Bandwidth = 0.04347

Econometria bayesiana (XIX)

Resultados para o modelo com *priori Cauchy*:

```
##
## Iterations = 5001:55000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 50000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                 Mean      SD  Naive SE Time-series SE
## (Intercept) -0.42520 0.68317 0.0030552      0.0113878
## age         -0.04987 0.02812 0.0001258      0.0004924
## raceblack    0.69043 0.45442 0.0020322      0.0083937
## raceother    0.77778 0.37570 0.0016802      0.0069519
## smoke        0.89363 0.35700 0.0015965      0.0068431
##
## 2. Quantiles for each variable:
##
##                 2.5%      25%      50%      75%    97.5%
## (Intercept) -1.90407 -0.84763 -0.39052  0.03433 0.827555
## age         -0.10276 -0.06886 -0.05029 -0.03219 0.008692
## raceblack   -0.16579  0.37846  0.67502  0.98608 1.624051
## raceother    0.07031  0.52314  0.76928  1.02059 1.533091
## smoke        0.22076  0.64436  0.88249  1.13209 1.613151
```

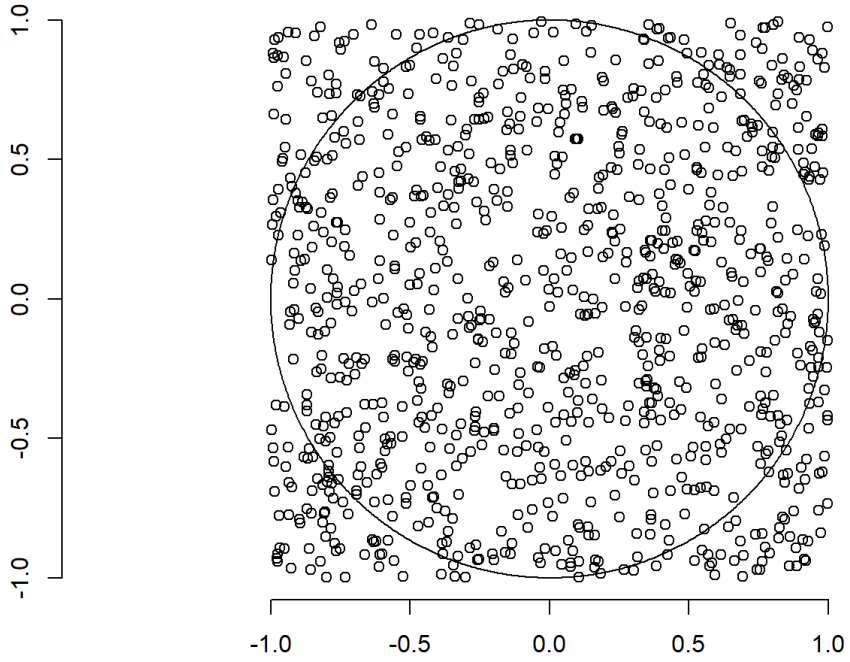
Métodos de Monte Carlo (I)

Os métodos de *Monte Carlo* são métodos numéricos (probabilísticos) para aproximar o cálculo de integrais multidimensionais. Por exemplo, imagine que é necessário calcular a área de um círculo unitário. Sabe-se que tal área é $A = \pi r^2$. Assim, no círculo unitário a área é $A = 3.1415$.

Uma forma de se obter a mesma área é gerar um número aleatório de pontos no quadrado com vértices $(-1, -1)$, $(-1, 1)$, $(1, -1)$ e $(1, 1)$ e contar a quantidade de pontos dentro do círculo unitário centrado em $(0, 0)$.

O método foi inspirado no distrito de Monte Carlo (http://monte-carlo.mc/en) no principado de Mônaco; locais famosos por seus cassinos.

Métodos de Monte Carlo (II)



O quadrado tem área igual a 4. Dentro do círculo existem 757 pontos e dentro do retângulo existem 1000 pontos. A área do círculo é 0.757 a área do retângulo, ou seja, a área é 3.028. Este valor está próximo do calculado pela fórmula.

MCMC - Markov Chain Monte Carlo (I)

Os métodos **MCMC** (*métodos de Monte Carlo via cadeia de Markov*) são métodos de Monte Carlo implementados através do uso de *cadeias de Markov*. Dentre os principais métodos, cita-se:

- **Metropolis–Hastings:** usada para gerar observações de uma distribuição $\pi(\theta|X)$ partindo de uma geradora de candidados $g(\theta_1|\theta_0)$ mais simples. O método em seu estado estacionário gera uma sequência de observações compatíveis com a distribuição de interesse.

- **Gibbs sampling:** permite gerar observações da distribuição conjunta $g(X_1, X_2, \ldots, X_p)$ a partir das distribuições condicionais $g(X_1), g(X_2|X_1), \ldots, g(X_p|X_1, X_2, \ldots, X_{p-1})$. O método em seu estado estacionário fornece vetores $\boldsymbol{x}_k = [x_{k1}, x_{k2}, \ldots, x_{kp}]$ da distribuição conjunta desejada.

- **Reversible-jump:** Este método é uma variante do Metropolis–Hastings que permite que a dimensão do espaço dos candidatos seja variante.

MCMC - Markov Chain Monte Carlo (II)

O método **Metropolis-Hastings** consiste no seguinte:

- Partindo de um valor inicial θ_0, na k-ésima etapa, gera-se um candidato θ da distribuição geradora de candidatas $g(\theta|\theta_{k-1})$.
- Se $h(\theta)$ é a distribuição para a qual se deseja gerar observações, calcula-se a **probabilidade de aceitação**

$$A(\theta_{k-1}, \theta) = \min\left(1, \frac{h(\theta_{k-1})g(\theta|\theta_{k-1})}{h(\theta)g(\theta_{k-1}|\theta)}\right).$$

- De uma uniforme contínua em $[0, 1]$ gera-se um valor $u \in [0, 1]$ e se $u \leq A(\theta_{k-1}, \theta)$ aceita-se $\theta_k = \theta$.

MCMC - Markov Chain Monte Carlo (II)

O método **Gibbs sampling** consiste em, partindo de um chute inicial $\boldsymbol{x}_0 = [x_{01}, x_{02}, \ldots, x_{0p}]$, gerar no k-ésimo passo, x_{k1} da distribuição $g(X_1|X_2 = x_{k-1,2}, X_3 = x_{k-1,3}, \ldots, X_p = x_{k-1,p})$, x_{k2} da distribuição $g(X_2|X_3 = x_{k-1,3}, X_4 = x_{k-1,4}, \ldots, X_p = x_{k-1,p})$, sucessivamente, e x_{kp} da distribuição $g(X_p)$.

Quando a cadeia estiver estacionária, os vetores $\boldsymbol{x}_k = [x_{k1}, x_{k2}, \ldots, x_{kp}]$ gerados serão observações amostrais da distribuição conjunta $g(X_1, X_2, X_3, \ldots, X_p)$.

Modelo VAR (Vetor AutoRegressivo)

O modelo VAR (Vetor AutoRegressivo) é utilizado para ajustar simultaneamente um conjunto de séries temporais correlacionadas.

Tal modelo é definido no instante $t \geq 0$ como

$$\boldsymbol{y}_t = c + \sum_{j=i}^{p} \boldsymbol{A}_j \boldsymbol{y}_{t-j} + \epsilon_t$$

em que $\boldsymbol{y}_t$ é um vetor de dimensão k dos valores das séries no instante t, $\boldsymbol{A}_j$ é a matriz de coeficientes correlacionados ao j-ésimo lag de $\boldsymbol{y}_t$, c é uma constante e ϵ_t é um vetor de dimensão k de erros com média $\mathbf{0}$ e variância $\boldsymbol{\Sigma}$.

Modelo VAR Bayesiano (I)

O ajuste dos modelos VAR através de métodos bayesianos é feito com a aplicação do **Gibbs sampling**. Para demonstrar o método usaremos o conjunto de dados apresentado em "Lütkepohl, H. *New introduction to multiple time series analysis*. 2. ed. Berlin: Springer, 2007" das séries de investimento, renda e consumo na Alemanha Oriental entre primeiro trimestre de 1960 e quarto trimestre de 1982.

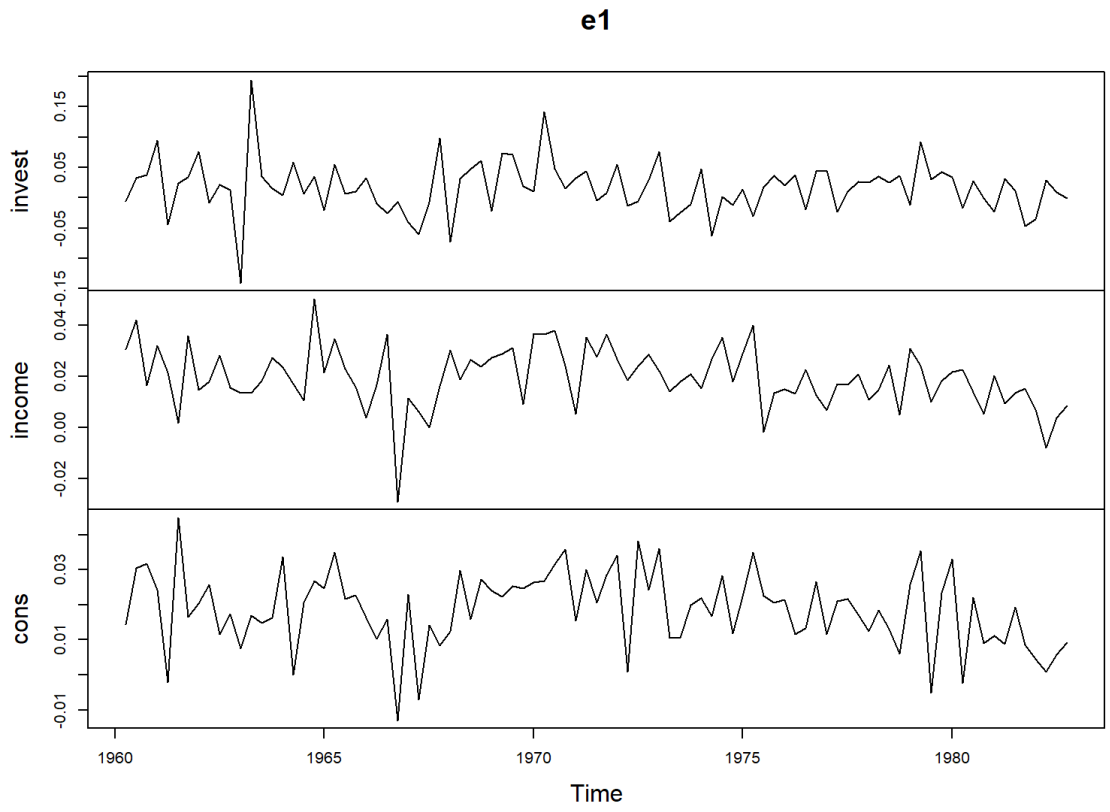
```
library(bvartools)
```

```
##
## Attaching package: 'bvartools'
```

```
## The following object is masked from 'package:coda':
##
##     thin
```

```
data("e1")          # Carrega os dados
e1 = diff(log(e1))  # Calcula o log das diferenças de 1a ordem
```

Modelo VAR Bayesiano (II)



Modelo VAR Bayesiano (III)

O primeiro passo é gerar os vetores $\boldsymbol{y}$ e $\boldsymbol{x}$ para ajustar o modelo VAR

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{u}_t$$

em que $\boldsymbol{u}_t$ segue uma distribuição $N(\boldsymbol{0}, \boldsymbol{\Sigma})$.

```
# Modelo VAR com 2 lags
d = gen_var(e1, p = 2, deterministic = "const")

y = d$Y
z = d$Z
```

Modelo VAR Bayesiano (IV)

Para estimar o VAR frequentista, calcula-se $\boldsymbol{yx}'(\boldsymbol{xx}')^{-1}$. O resultado servirá de base para avaliar a precisão do modelo bayesiano.

```
A_freq = tcrossprod(y, z) %*% solve(tcrossprod(z))
round(A_freq, 3)
```

```
##        invest.1 income.1 cons.1 invest.2 income.2 cons.2  const
## invest   -0.273    0.337  0.652   -0.134    0.183  0.598 -0.010
## income    0.043   -0.123  0.305    0.062    0.021  0.049  0.013
## cons      0.003    0.289 -0.285    0.050    0.366 -0.116  0.012
```

Modelo VAR Bayesiano (V)

A matriz de variância Σ é calculada por

```
u_freq <- y - A_freq %*% z
u_sigma_freq <- tcrossprod(u_freq) / (ncol(y) - nrow(z))
round(u_sigma_freq * 10^4, 2)
```

```
##        invest income cons
## invest  19.62   0.62 1.41
## income   0.62   1.26 0.64
## cons     1.41   0.64 0.99
```

Modelo VAR Bayesiano (VI)

```
iter <- 30000    # Número de iterações do Gibbs sampler
burnin <- 15000  # Número de iterações no burn-in
store <- iter - burnin

t <- ncol(y) # Número de observações
k <- nrow(y) # Número de observações no histórico.
m <- k * nrow(z) # Número de coeficientes para estimar

# Define prioris (não informativa)
a_mu_prior <- matrix(0, m) # Vetor de médias da priori
a_v_i_prior <- diag(0, m)  # Inversa da matriz de covariância da priori

u_sigma_df_prior <- 0                   # Graus de liberdade a priori
u_sigma_scale_prior <- diag(0, k)       # Matriz de covariancia a priori
u_sigma_df_post <- t + u_sigma_df_prior # Graus de liberdade a posteriori
```

Modelo VAR Bayesiano (VII)

```r
# Valores iniciais
u_sigma_i <- diag(.00001, k)
u_sigma <- solve(u_sigma_i)

# Container de dados para as simulação
draws_a <- matrix(NA, m, store)
draws_sigma <- matrix(NA, k^2, store)

# Amostrator de Gibbs (Gibbs sampler)
for (draw in 1:iter) {
  # Amostra valores da posteriori (média)
  a <- post_normal(y, z, u_sigma_i, a_mu_prior, a_v_i_prior)

  # Amostra valores da posteriori (covariância)
  u <- y - matrix(a, k) %*% z    # Obtém os resíduos
  u_sigma_scale_post <- solve(u_sigma_scale_prior + tcrossprod(u))
  u_sigma_i <- matrix(rWishart(1, u_sigma_df_post, u_sigma_scale_post)[,, 1], k)
  u_sigma <- solve(u_sigma_i)    # Inverte Sigma_i para obter Sigma

  # Armazena os resultados
  if (draw > burnin) {
    draws_a[, draw - burnin] <- a
    draws_sigma[, draw - burnin] <- u_sigma
  }
}
```

Modelo VAR Bayesiano (VII)

Para obter os coeficientes:

```
A <- rowMeans(draws_a)   # Obtém as médias para cada linha
A <- matrix(A, k)        # Transforma os vetores de média em matriz
A <- round(A, 3)         # Arredonda os valores
dimnames(A) <- list(dimnames(y)[[1]], dimnames(x)[[1]]) # Renomeia as dimensões

A # Exibe
```

```
##          [,1]   [,2]   [,3]   [,4]  [,5]   [,6]   [,7]
## invest -0.272  0.337  0.656 -0.134 0.183  0.595 -0.010
## income  0.043 -0.122  0.304  0.062 0.021  0.048  0.013
## cons    0.003  0.290 -0.285  0.050 0.367 -0.117  0.012
```

Modelo VAR Bayesiano (VIII)

Para obter as covariâncias:

```
Sigma <- rowMeans(draws_sigma)  # Obtém as médias para cada linha
Sigma <- matrix(Sigma, k)       # Transforma o vetor de média em matriz
Sigma <- round(Sigma * 10^4, 2) # Arredonda os valores
dimnames(Sigma) <- list(dimnames(y)[[1]], dimnames(y)[[1]]) # Renomeia as dimensões

Sigma # Exibe
```

```
##        invest income cons
## invest  20.70   0.65 1.48
## income   0.65   1.32 0.67
## cons     1.48   0.67 1.04
```

Modelo VAR Bayesiano (IX)

Para fazer previsões das séries e para diagnósticos é necessário transformar o objeto criado para um objeto `var`.

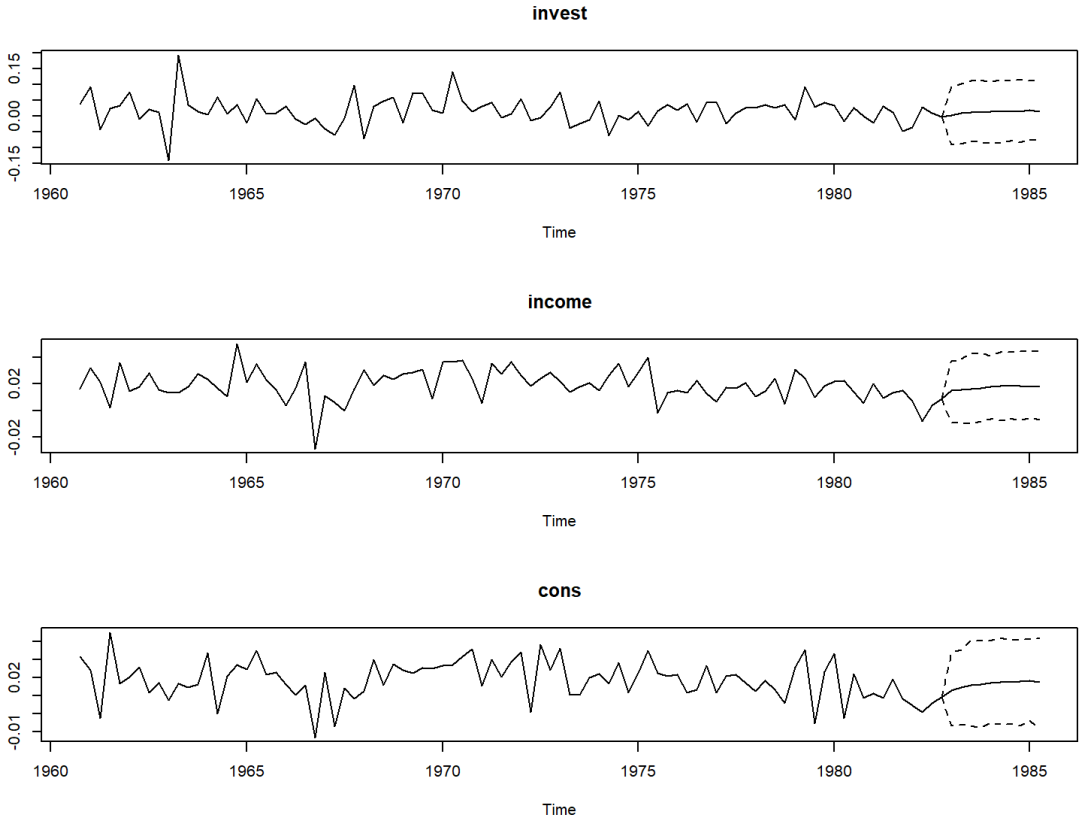
Transforma o resultado em um objeto VAR.

```
bvar_est <- bvar(y = y, x = z, A = draws_a[1:18,], C = draws_a[19:21, ],
                 Sigma = draws_sigma)
bvar_est <- thin(bvar_est, thin = 15)
```

Modelo VAR Bayesiano (X)

É possível fazer previsões do modelo:

```
bvar_pred <- predict(bvar_est, n.ahead = 10, new_D = rep(1, 10))
plot(bvar_pred)
```



Modelo VAR Bayesiano (XI)

"As funções de impulso-resposta são usadas para descrever a reação (da economia) aos impulsos nas variáveis exogenas ao longo do tempo, o que a economia chama choques." (*Wikipedia - Impulse response function.*)

A *resposta-impulso do erro de previsão* é definida como

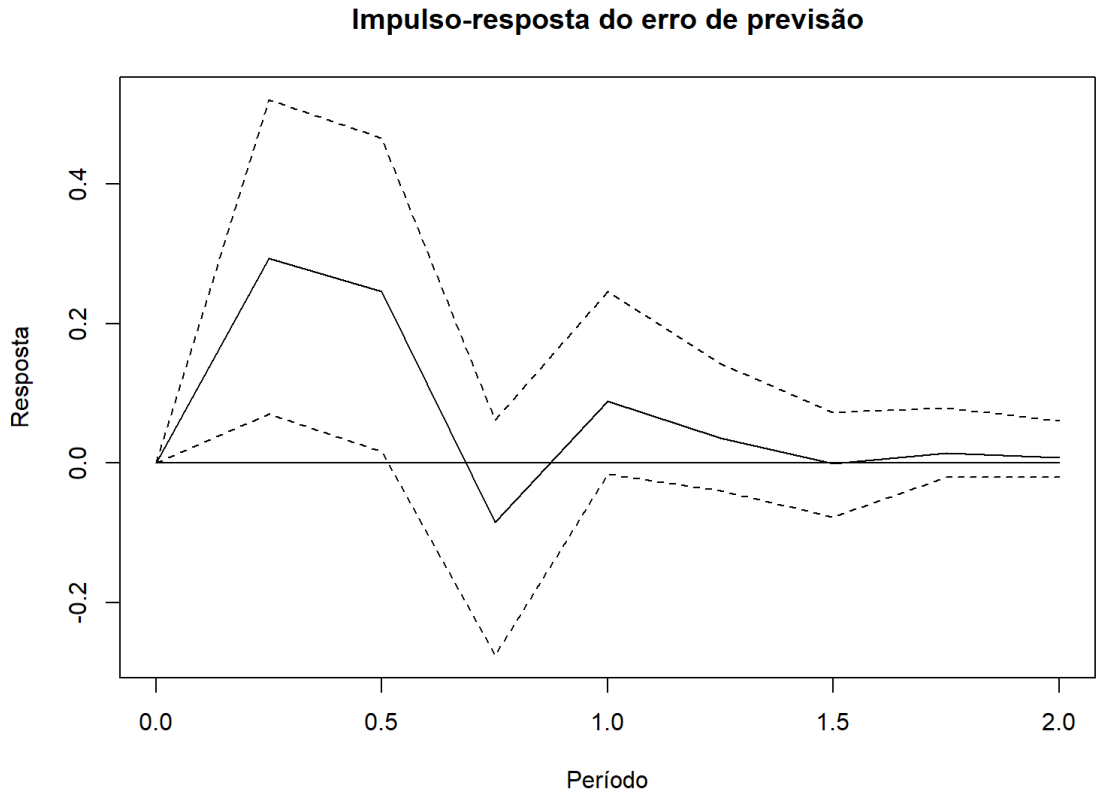
$$\boldsymbol{\phi}_k = \sum_{j=1}^{k} \boldsymbol{\phi}_{k-j} \boldsymbol{A}_j,\ k = 1, 2, \ldots, p$$

Impulso-resposta ortogonalizado são calculados como $\boldsymbol{\theta}_i^o = \boldsymbol{\Phi}_i \boldsymbol{P}$, em que $\boldsymbol{P}$ é a matriz triangular inferior da decomposição de Choleski $\boldsymbol{\Sigma}$. $\boldsymbol{A}_0$ é assumido ser a matriz identidade.

Impulso-resposta (estrutural) generalizado para a série j são calculados como $\boldsymbol{\theta}_{ji}^g = \sigma_{jj}^{-1/2} \boldsymbol{\phi}_i \boldsymbol{A}_0^{-1} \boldsymbol{\Sigma} e_j$, em que σ_{jj} é a variância do j-ésimo elemento diagonal de $\boldsymbol{\Sigma}$ e e_i é um vetor de seleção contendo o valor 1 no j-ésimo elemento e 0 nos demais. A matriz $\boldsymbol{A}_0$, se não fornecida, é assumida ser a identidade.

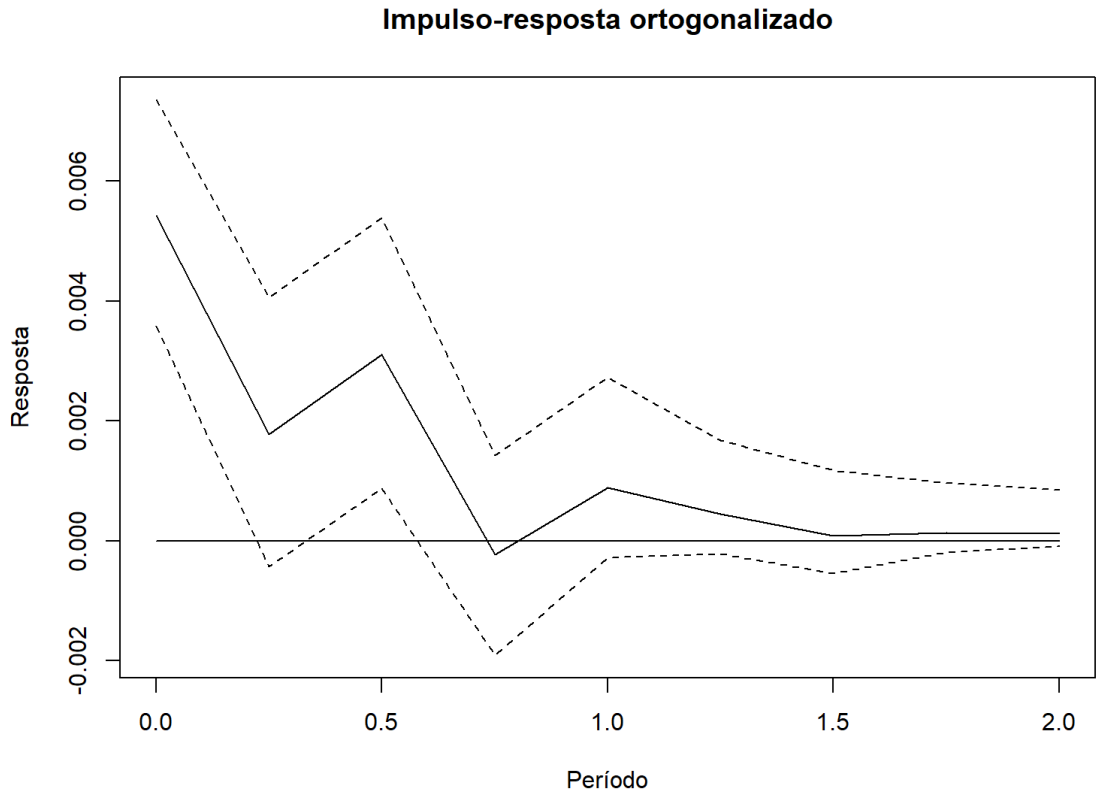
Modelo VAR Bayesiano (XII)

```
FEIR <- irf(bvar_est, impulse = "income", response = "cons", n.ahead = 8)
plot(FEIR, main = "Impulso-resposta do erro de previsão",
     xlab = "Período", ylab = "Resposta")
```



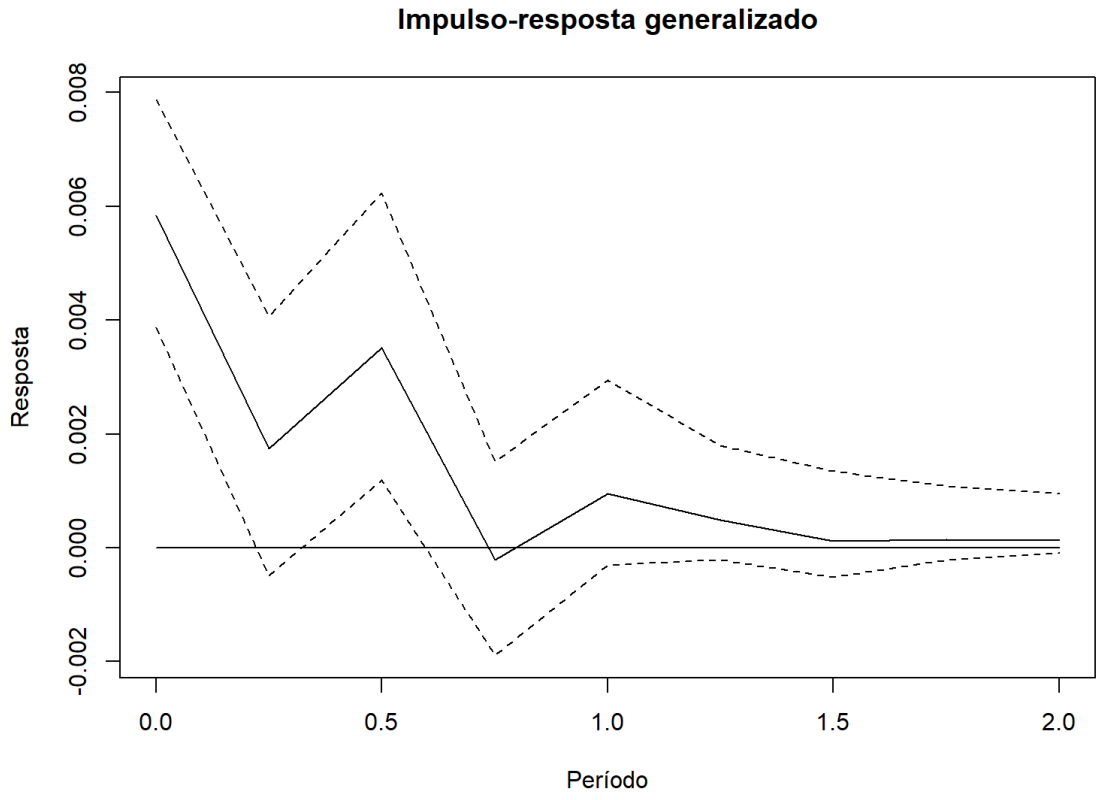
Modelo VAR Bayesiano (XIII)

```r
OIR <- irf(bvar_est, impulse = "income", response = "cons", n.ahead = 8, type = "oir")
plot(OIR, main = "Impulso-resposta ortogonalizado",
     xlab = "Período", ylab = "Resposta")
```



Modelo VAR Bayesiano (XIV)

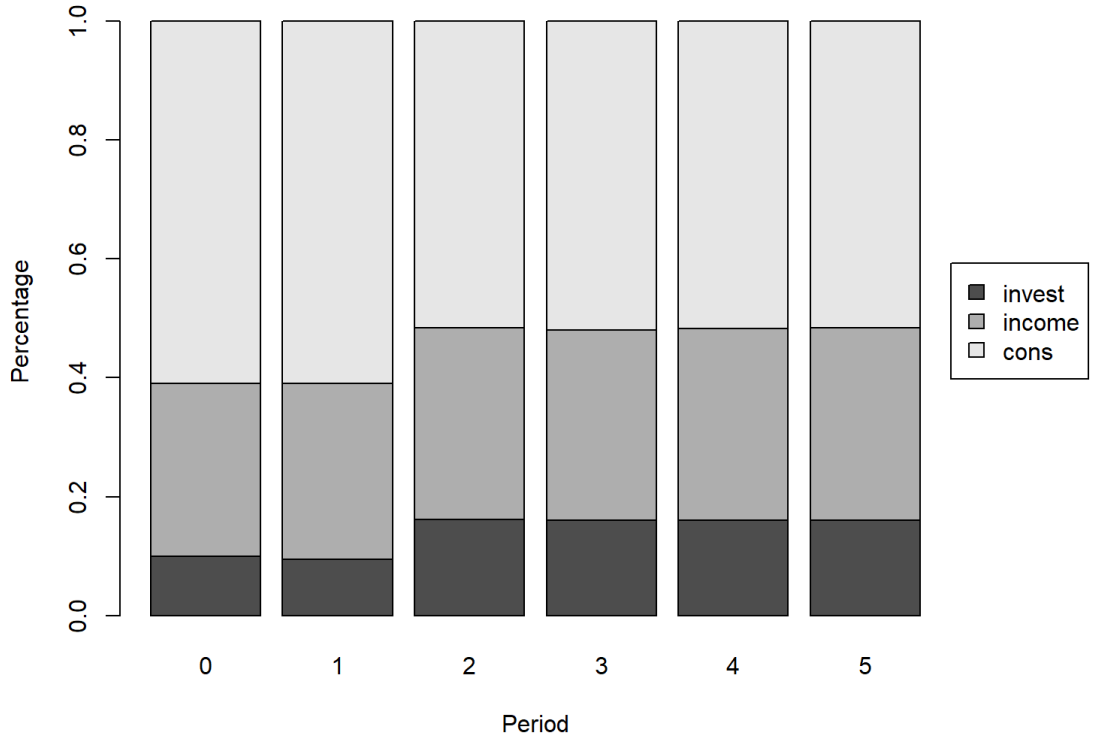
```r
GIR <- irf(bvar_est, impulse = "income", response = "cons", n.ahead = 8, type = "gir")
plot(GIR, main = "Impulso-resposta generalizado",
     xlab = "Período", ylab = "Resposta")
```



Modelo VAR Bayesiano (XV)

É possível obter a decomposição da variância do erro de previsão com relação ao impulso-resposta ortogonalizado:

```
bvar_fevd_oir <- fevd(bvar_est, response = "cons")
plot(bvar_fevd_oir, main = "OIR-based FEVD of consumption")
```



Modelo VAR Bayesiano (XVI)

É possível obter a decomposição da variância do erro de previsão com relação ao impulso-resposta generalizado:

```
bvar_fevd_gir <- fevd(bvar_est, response = "cons", type = "gir")
plot(bvar_fevd_gir, main = "GIR-based FEVD of consumption")
```