

# Métodos analíticos e numéricos para a obtenção de distribuições à posteriori: exemplos multivariados

Prof. Caio Azevedo

# Motivação

- Objetivo: calcular aproximações analíticas e/ou numéricas para integrais multivariadas (k-variadas).
- Relembrando se  $f(., .)$  for uma densidade bivariada, então:

$$\int_a^b \left( \int_c^d f(x, y) dx \right) dy = F(b, d) - F(a, d) - F(b, c) + F(a, c),$$

em que

$$f(x, y) = \frac{\partial^2}{\partial x \partial y'} f(x, y) = \frac{\partial^2}{\partial y \partial x'} F(x, y).$$

# Abordagens

- Algumas opções:
  - Aproximações analíticas (Laplace) e INLA.
  - Aproximações numéricas não estocásticas (Quadratura, Quadratura adaptativa).
  - Aproximações numéricas estocásticas (não-iterativas) (Monte Carlo).
  - Aproximações numéricas estocásticas iterativas (Monte Carlo via Cadeias de Markov).
  - Veja referências [aqui](#).
- Alguns resultados vistos para o caso univariado ([aqui](#)) são úteis para desenvolver/demonstrar alguns resultados no caso multivariado.

## Aproximações analíticas

- Sejam  $\mathbf{x} = (x_1, \dots, x_k)' \in \mathcal{R}^k$  e  $\mathbf{x}_0 = (x_{01}, \dots, x_{0k})' \in \mathcal{R}^k$ .
- Vamos assumir, sob certas condições, para um dado  $\mathbf{x}_0$ , que:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)' \left. \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) + R,$$

em que  $R$  é suficientemente desprezível.

- Desejamos calcular (em que os limites podem ser  $\pm\infty$ )

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} e^{Mf(\mathbf{x})} d\mathbf{x}, \mathbf{x} = (x_1, \dots, x_k)'.$$

- Suponha ainda que existe um único máximo global para  $f(\cdot)$ .
- Defina  $A = (a_1, b_1) \times (a_2, b_2) \times \dots \times (a_k, b_k)$ .

# Aproximação de Laplace (AL)

- Seja  $\mathbf{x}_0$  o máximo global de  $f(\cdot)$ . Então, de forma semelhante ao caso univariado, temos que :

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)' \left. \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0)$$

- AL (em que  $\mathbf{X} \sim N_k(\mathbf{x}_0, \left( \left. \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\mathbf{x}_0} \right)^{-1})$ ):

$$\begin{aligned} & \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} e^{Mf(\mathbf{x})} d\mathbf{x} \approx \\ & e^{Mf(\mathbf{x}_0)} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} e^{-M \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)' \left[ - \left. \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\mathbf{x}_0} \right] (\mathbf{x} - \mathbf{x}_0)} d\mathbf{x} \\ & = e^{Mf(\mathbf{x}_0)} (2\pi)^{k/2} \left| \left( \left. \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{\mathbf{x}=\mathbf{x}_0} \right)^{-1} \right|^{1/2} \times P(\mathbf{X} \in A). \end{aligned}$$

# Características da AL

- Relativamente rápida.
- Adequada quando o integrando é unimodal e o máximo pode ser obtido facilmente (ainda que seja necessário utilizar métodos numéricos).
- Inadequada: em integrais múltiplas à medida que a dimensão aumenta e/ou o máximo é complicado de ser obtido.
- Mesmo no caso univariado, não é adequada para funções complicadas: vários máximos locais, assimetria, etc.

## Exemplo: Dirichlet bidimensional

- A distribuição Dirichlet é uma generalização da distribuição beta ([aqui](#) e [aqui](#)).
- Dizemos que  $\mathbf{X} = (X, Y)' \sim \text{Dirichlet}(2, \alpha, \beta, \gamma)$  se:

$$\begin{aligned} f(x, y) &= \frac{1}{\beta(\alpha, \beta, \gamma)} x^{\alpha-1} y^{\beta-1} (1-x-y)^{\gamma-1} \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,1-x)}(y) \\ &= \frac{1}{\beta(\alpha, \beta, \gamma)} x^{\alpha-1} y^{\beta-1} (1-x-y)^{\gamma-1} \mathbb{1}_{(0,1-y)}(x) \mathbb{1}_{(0,1)}(y) \end{aligned}$$

em que  $\beta(\alpha, \beta, \gamma) = \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\alpha + \beta + \gamma)}$ .

- Função de interesse:

$$F_{(X, Y)}(1/3, 1/3) = \int_0^{1/3} \left( \int_0^{1/3} f(x, y) dy \right) dx.$$

## Exemplo: Dirichlet bidimensional

- Pode-se provar (exercício) que (ponto de máximo) é dado por  $x_0 = \frac{\alpha - 1}{\alpha + \beta + \gamma - 3}$  e  $y_0 = \frac{\beta - 1}{\alpha + \beta + \gamma - 3}$ .
- Além disso

$$\begin{aligned} -\Sigma^{-1} &= \frac{\partial}{\partial \mathbf{x} \partial \mathbf{x}'} f(\mathbf{x}) \\ &= \begin{bmatrix} -\frac{\alpha - 1}{x^2} - \frac{\gamma - 1}{(1 - x - y)^2} & -\frac{\gamma - 1}{(1 - x - y)^2} \\ -\frac{\gamma - 1}{(1 - x - y)^2} & -\frac{\beta - 1}{y^2} - \frac{\gamma - 1}{(1 - x - y)^2} \end{bmatrix} \end{aligned}$$



## Cont.

- $f(\mathbf{x}_0) = \exp((\alpha - 1) \ln(x) + (\beta - 1) \ln(y) + (\gamma - 1) \ln(1 - x - y))$ .
- $h \approx \frac{1}{\beta(\alpha, \beta, \gamma)} e^{f(\mathbf{x}_0)} 2\pi (|\boldsymbol{\Sigma}|)^{1/2} P(0 < X_1 < 1/3, 0 < X_2 < 1/3)$ ,  
 $\mathbf{X} \sim N((x_0, y_0), \boldsymbol{\Sigma}_{\mathbf{x}=(x_0, y_0)'})$ .
- Vamos considerar  $\alpha = \beta = \gamma = 2$ .
- Vamos comparar o valor de  $h$  com uma aproximação empírica, baseada na fda empírica da distribuição de Dirichlet (chamemo-na de  $g$ ), usando a função `rdirichlet` do pacote `LaplaceDemon`.
- Ou sejam vamos gerar  $m$  valores de uma distribuição  $Dirichlet(2, 2, 2, 2)$  e calcular o número de vezes  $(x_i < 1/3, y_i < 1/3)$ ,  $i=1, 2, \dots, m$ .

## Aproximação da Laplace

```
alphap <- 2; betap <- 2; gammap <- 2
x0 <- (alphap-1)/(alphap+gammap+betap -3)
y0 <- (betap-1)/(alphap +gammap +betap-3)
fx0 <- log(x0^(alphap-1)*y0^(betap-1)*(1-x0-y0)^(gammap-1))
Hxx <- -(alphap-1)/(x0^2) - (gammap-1)/(1-x0-y0)^2
Hyy <- -(betap-1)/(y0^2) - (gammap-1)/(1-x0-y0)^2
Hxy <- -(gammap-1)/(1-x0-y0)^2
m.H <- rbind(cbind(Hxx,Hxy),cbind(Hxy,Hyy))
m.cov <- -solve(m.H)
```

## Aproximação da Laplace (Cont.)

```
F1 <- pmnorm(c(1/3,1/3),c(x0,y0),(m.cov))
F2 <- pmnorm(c(0,1/3),c(x0,y0),(m.cov))
F3 <- pmnorm(c(1/3,0),c(x0,y0),(m.cov))
F4<- pmnorm(c(0,0),c(x0,y0),(m.cov))
ALint <- (gamma(alphap+gammap+betap)/(gamma(alphap)*
gamma(betap)*gamma(gammap)))*exp(fx0)*(2*pi)^(2/2)*
det(m.cov)^(1/2)*(F1-F2-F3+F4)
ALint
```

# FDA empírica

```
m<-10000  
m_dirch <- rdirichlet(m, c(alphap,betap,gammap))  
Fbd<- mean(as.numeric(m_dirch[,1]<=1/3 & m_dirch[,2]<=1/3))  
Fbd
```

# Integração por Quadratura

- Substituir (no caso multivariado) o cálculo da área sob a superfície pela soma das áreas de um número finito de paralelepípedos.
- Ponto-chave: Definir os pontos e pesos de quadratura.
- Formas de cálculo dos pontos e pesos:
  - Não-adaptativa: mantém-se fixo os pontos e pesos.
  - Adaptativa: muda-se os pontos e pesos.
- As questões relevantes neste contexto são semelhantes àquelas para o caso univariado.
- No entanto, o número de conjunto de pontos de quadraturas tende a crescer muito com o aumento do número da dimensão das integrais de interesse.

# Características

- Em geral, na Estatística, deseja-se calcular

$$\mathcal{E}(w(\mathbf{X})) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} w(\mathbf{x})f(\mathbf{x})d\mathbf{x},$$

em que  $f(\cdot)$  é alguma fdp  $k$ -variada ( $\mathbf{X} = (X_1, X_2, \dots, X_k)'$ ).

- Assim,  $f(\cdot)$  terá massa relevante em apenas um subconjunto de  $\mathcal{R}^k$ .
- Define-se conjunto de pontos,  $(x_{11}, x_{12}, \dots, x_{1k}); \dots; (x_{m1}, x_{m2}, \dots, x_{mk})$ , em que  $x_{ij}$  é o  $i$ -ésimo ponto de quadratura associado à  $j$ -ésima dimensão,  $i = 1, 2, \dots, m$  e  $j = 1, 2, \dots, k$ , e os respectivos pesos associados  $A_1, A_2, \dots, A_k$ ,  $A = \prod_{j=1}^k A_j$ ,  $A_j = A(x_{ij})$ ,  $i = 1, 2, \dots, m$ .
- Comumente assume-se a mesma quantidade de pontos de quadratura, para cada dimensão.

# Características

- Em geral, os pesos correspondem aos comprimentos dos intervalos determinados pelos pontos de quadraturas, em cada dimensão.
- Pode-se imaginar que cada ponto corresponde ao valor médio de intervalos de mesmo comprimento.
- Pontos de Quadratura:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mk} \end{bmatrix}_{m \times k}$$

# Continuação

- À semelhança do que ocorre no caso univariado, tem-se o mesmo comprimento de intervalos, ao longo das dimensões.
- Com efeito, comprimentos dos intervalos de Quadratura:

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_k \\ A_1 & A_2 & \dots & A_k \\ \vdots & \vdots & \ddots & \vdots \\ A_1 & A_2 & \dots & A_k \end{bmatrix}_{m \times k}$$

- Assim, tem-se um vetor coluna com todos os elementos iguais à

$$A = \prod_{j=1}^k A_j.$$



# Continuação

- Faz-se necessário a construção de uma matriz com todas as  $k$ -uplas possíveis de serem formadas com os pontos de Quadratura de cada dimensão.
- Pelo princípio fundamental da contagem temos um total de  $m^k$   $k$ -uplas possíveis.
- Analogamente, constrói-se um vetor com elementos todos iguais à  $A$ , de dimensão  $m^k \times 1$ .

# Continuação

- Exemplo  $m = 3, k = 2$

$$\mathbf{X}^* = \begin{bmatrix} x_{11} & x_{12} \\ x_{11} & x_{22} \\ x_{11} & x_{32} \\ x_{21} & x_{12} \\ x_{21} & x_{22} \\ x_{21} & x_{32} \\ x_{31} & x_{12} \\ x_{31} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} m^k \times k$$

# Continuação

- Exemplo  $m = 3, k = 2$

$$\mathbf{A}^* = \begin{bmatrix} A \\ A \\ A \\ A \\ A \\ A \\ A \\ A \\ A \\ A \end{bmatrix}_{m^k \times 1}$$

## Cont.

- Seja  $\mathcal{I}$  o conjunto que contém todas as  $k$ -uplas necessárias.
- Assim

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} w(\mathbf{x})f(\mathbf{x})d\mathbf{x} \approx \sum_{i \in \mathcal{I}} A_i w(\mathbf{x}_i.)f(\mathbf{x}_i.)$$

em que  $\mathbf{x}_i.$  é o vetor que contém a  $k$ -ésima upla.

## Cont.

- Existem várias formas de se determinar os pontos e pesos. Em geral, faz-se tal escolha, em torno do máximo (moda) de  $f(\cdot)$ .
- Pontos ótimos, em geral, são obtidos através de certos polinômios.
- Polinômios de Gauss-Hermite, Gauss-Legendre, Jacobi, Legendre, Gauss-Laguerre etc. [Abramowitz & Stegun \(1972\)](#).

## Exemplo: Normal trivariada

- Seja  $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , em que  $\boldsymbol{\mu} = (-1, 0, 2)'$  e

$$\boldsymbol{\Sigma} = \begin{bmatrix} 5 & 1 & -2 \\ 1 & 6 & 3 \\ -2 & 3 & 8 \end{bmatrix}.$$

- Vamos calcular  $I = \int_{\mathcal{R}^3} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{x}$ . Sabemos que  $I = 1$ .
- Usaremos  $m = 10$  pontos de quadratura por dimensão.
- Abrir o arquivo no R. Resultado = 0,9999862

# Quadratura Adaptativa

- Atualização dos pontos e pesos de quadratura.
- Nosso interesse continua sendo

$$h = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} w(\mathbf{x})f(\mathbf{x})d\mathbf{x}.$$

- Algoritmo

- Calcule duas aproximações para h (considerando k=2),

$$S_1 = S(f(x); \mathbf{a}, \mathbf{b}) = \sum_{i=1}^m A_i w(\mathbf{x}_{i.}) f(\mathbf{x}_{i.}) \text{ e}$$

$$S_2 = S(f(x); a_1, (a_1 + b_1)/2, a_2, (a_2 + b_2)/2) + S(f(x); (a_1 + b_1)/2, b_1, (a_2 + b_2)/2, b_2) + S(f(x); (a_1 + b_1)/2, b_1, a_2, (a_2 + b_2)/2) + S(f(x); (a_1 + b_1)/2, b_1, (a_2 + b_2)/2, b_2).$$

- Calcule  $\epsilon = |S_1 - S_2|$ . Se  $\epsilon < \tau$ , pare, caso contrário, aplique o passo 1, recursivamente, até que a precisão requerida seja atingida.

# Quadratura Adaptativa no R

- Pacote  **cubature** .
- Função  **adaptIntegrate**  .
- Permite calcular integrais, usando a quadratura adaptativa, para funções pré-definidas pelo usuário, em um dado intervalo.
- Seja  $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , em que  $\boldsymbol{\mu} = (1/2, 1/2)'$  e
$$\boldsymbol{\Sigma} = \begin{bmatrix} 0,01 & 0 \\ 0 & 0,01 \end{bmatrix}.$$
- Vamos calcular  $I = \int_{\mathcal{R}^2} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{x}$ . Sabemos que  $I = 1$ .  
Resultado  $I = 1,000006$ .



# Adaptativa

```
M_2_SQRTPI <- 2/sqrt(pi)
a = 0.1
int.NM <- function(x) {
  s = sum((x-0.5)^2)
  (M_2_SQRTPI / (2. * a))^length(x) * exp (-s / (a * a))
}
adaptIntegrate(int.NM, rep(-Inf,2), rep(Inf,2), tol=1e-4)
```

# Integração por Monte Carlo

- Lei Forte dos Grandes Números: Seja  $\mathbf{X}_1, \mathbf{X}_2, \dots$  uma sequência iid, tal que  $\mathcal{E}(\mathbf{X}_i|\boldsymbol{\mu}) = \boldsymbol{\mu}, \forall i$ . Defina  $Y_1, Y_2, \dots$ ,  $Y_i = g(\mathbf{X}_i), i = 1, 2, \dots, \mathcal{E}(Y_i) = \mathcal{E}(g(\mathbf{X}_i)|\boldsymbol{\mu}) = h(\boldsymbol{\mu}) = \mu_y$ . Então

$$\bar{Y}_n = \sum_{i=1}^n \frac{1}{n} Y_i \xrightarrow{q.c.} \mu_y$$

- Integração por Monte Carlo.
  - 1 Gere um conjunto de  $m$  vetores aleatórios i.i.d, digamos

$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ .

- 2 Calcule  $h \approx \frac{1}{m} \sum_{i=1}^m w(\mathbf{x}_i)$ .

# Integração por Amostragem por Importância

- Muito esforço para simular pontos com massa desprezível.
- Pode ser complicado definir o valor esperado e/ou a distribuição apropriadas.
- Refinar a geração dos números aleatórios: método da amostragem por importância.

# Algoritmo

- Gere uma amostra aleatória,  $\mathbf{X}_1, \dots, \mathbf{X}_m$  de uma densidade candidata  $g(\cdot)$  com o mesmo suporte  $f(\cdot)$ .
- Assim

$$h = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} w(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_k}^{b_k} w(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x}.$$

- Construa pesos de importância

$$W(X_i) = \frac{f(X_i)}{g(X_i)}$$

- Calcule  $h \approx \frac{1}{m} \sum_{i=1}^m W(X_i) w(X_i)$

Exemplo:  $f(x_1, x_2) = e^{-90(x_1-0,5)^2} e^{-10(x_2+0,1)^4}$

- $h = \int_{-1}^1 \left( \int_{-1}^1 e^{-90(x_1-0,5)^2} e^{-10(x_2+0,1)^4} dx_1 \right) dx_2$ .
- Primeira abordagem: Integração por Monte Carlo: simular duas variáveis aleatórias, independente, Uniformes no intervalo  $[-1, 1]$  e calcular  $h \approx \frac{4}{m} \sum_{i=1}^m f(x_{i1}, x_{i2})$ .
- API: simular de duas variáveis aleatórias independentes normais truncadas no intervalo  $[-1, 1]$ .

$$g(x_1, x_2) \propto e^{-90(x_1-0,5)^2} e^{-10(x_2+0,1)^2}$$

# Comparação: API x MC

