Integração Numérica: Funções multivariadas

Prof. Caio Azevedo

- Objetivo: calcular aproximações analíticas numérica para integrais multivariadas (k-variadas).
- Relembrando: $\int_a^b \left(\int_c^d f(x,y)dx \right) dy = F(b,d) - F(a,d) - F(b,c) + F(a,c), F(x,y) = \frac{\partial^2}{\partial x \partial y'} f(x,y) = \frac{\partial^2}{\partial y \partial x'} f(x,y)$.
- Soluções:
  - Aproximações analíticas (Laplace).
  - Aproximações numéricas não estocásticas (Quadratura, Quadratura adaptativa).
  - Aproximações numéricas estocásticas (não-iterativas) (Monte Carlo).
  - Aproximações numéricas estocásticas iterativas (Monte Carlo via Cadeias de Markov).

Aproximações analíticas

- Sob certas condições, para um dado $\mathbf{x}_0$,

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

- Suponha que desejamos calcular

  $\int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_k}^{b_k} e^{Mf(\mathbf{x})} d\mathbf{x}, \mathbf{x} = (x_1, ..., x_k)$. Suponha que existe um único máximo global para $f(.)$.

Aproximação de Laplace (AL)

- Seja $\mathbf{x}_0$ o máximo global de $f(.)$. Então :

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)' \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

Cont.

- Aproximação de Laplace:

$$
\begin{aligned}
& \int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_k}^{b_k} e^{Mf(\mathbf{x})}d\mathbf{x} \approx \\
& e^{Mf(\mathbf{x}_0)}\int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_k}^{b_k} e^{-M\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)'\left[-\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}'}f(\mathbf{x}_0)\right](\mathbf{x}-\mathbf{x}_0)}d\mathbf{x} \\
= \quad & e^{Mf(x_0)}\left(\sqrt{2\pi M\left|-\frac{\partial^2}{\partial\mathbf{x}\partial\mathbf{x}'}f(\mathbf{x}_0)\right|}\right)^{-1} \\
\times \quad & P(\mathbf{X}\in A)
\end{aligned}
$$

$\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e A é a região de integração correspondente.

Característica da AL

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

Exemplo: Dirichlet

- $f(x,y) = \frac{1}{\beta(\alpha,\beta,\gamma)} x^{\alpha-1} y^{\beta-1} (1-x-y)^{\gamma-1} \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,1-x)}(y) = \frac{1}{\beta(\alpha,\beta,\gamma)} x^{\alpha-1} y^{\beta-1} (1-x-y)^{\gamma-1} \mathbb{1}_{(0,1-y)}(x) \mathbb{1}_{(0,1)}(y)$; $\beta(\alpha,\beta,\gamma) = \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}{\gamma(\alpha+\boldsymbol{\beta}+\gamma)}$.
- Função: $f(x) = \int_0^{1/3} \left(\int_0^{1/3} f(x,y) dy \right) dx$
- Pode-se provar que $x_0 = \frac{\alpha-1}{\alpha+\beta+\gamma-3}$ e $y_0 = \frac{\beta-1}{\alpha+\beta+\gamma-3}$.
- Além disso

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

Cont.

- $f(\mathbf{x}_0) = exp((\alpha - 1)\ln(x) + (\beta - 1)\ln(y) + (\gamma - 1)\ln(1 - x - y)))$.
- $h \approx \frac{1}{\boldsymbol{\beta}(\alpha,\beta,\gamma)} e^{f(x_0)} \left(2\pi|\boldsymbol{\Sigma}|\right)^{-1} P(0 < X_1 < 1/3, 0 < X_2 < 1/3)$,
  $\mathbf{X} \sim N((x_0, y_0), \boldsymbol{\Sigma}(x_0, y_0))$.

Integração por Quadratura

- Substituir (no caso multivarido) o cálculo da área sob a curva 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: mantem-se fixo os pontos e pesos.
  - Adaptativa: muda-se os pontos e pesos.

Características

- Em geral, na Estatística, deseja-se calcular $\mathcal{E}(w(\mathbf{X})) = \int_{a_1}^{b_1}\int_{a_2}^{b_2} \ldots \int_{a_k}^{b_k} w(\mathbf{x})f(\mathbf{x})d\mathbf{x}$, onde $f(.)$ é alguma fdp.
- Assim, $f(.)$ terá massa relevante em apenas um subconj. de $\mathcal{R}$.
- Define-se um conjunto de pontos, $(x_{11}, x_{12}, ..., x_{1k})$; $(x_{21}, x_{22}, ..., x_{2k})$; ...; $(x_{m1}, x_{m2}, ..., x_{mk})$ e os respectivos pesos associados $A_1, A_2, ..., A_k, A = \prod_{j=1}^{k} A_j, A_j = A(x_{ij}), i = 1, 2, ..., m,$. Em geral, os pesos correspondem aos comprimentos dos intervalores determinados pelos pontos de quadraturas, em cada dimensão.
- Pode-se imaginar que cada ponto corresponde ao valor médio de intervalos de mesmo comprimento.

Continuação

- Pontos de Quadratura:

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

Continuação

- À semelhança do que ocorre no caso univariado, tem-se o mesmo comprimento para cada dimensão, ou seja
- Comprimentos dos intervaloes 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.

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

- Analogamente, constroi-se um vetor com elementos todos igaus à A, de dimensão $m^k \times 1$

Continuação

- Exemplo $m = 3, k = 2$

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

Cont.

- Seja $\mathcal{I}$ o conjunto que contêm todas a 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 contem à k ésima upla.

Cont.

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

Exemplo: Normal multivariada

- Seja $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Ver programa no arquivo para R.

Quadratura Adptativa

- Atualização dos pontos e pesos de quadratura.
- Nosso interesse continua sendo $h = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \ldots \int_{a_k}^{b_k} w(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$.
- Algoritmo
  - Calcule duas aproximações para h (considerando k=2), $S_1 = S(f(x); \mathbf{a}, \mathbf{b}) = \sum_{i=1}^{m} A_i w(\mathbf{x}_{i.}) f(\mathbf{x}_{i.})$ 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 Adptativa no R

- Pacote **cubature**.
- Funação **adaptIntegrate** .
- Permite calcular integrais, usando a quadratura adaptativa, para funções pré-definidas pelo usuário, em um dado intervalo.

Integração por Monte Carlo

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

$$\overline{Y}_n = \sum_{i=1}^{n} \frac{1}{n} Y_i \overset{q.c.}{\rightarrow} \mu_y$$

- Integração por Monte Carlo.

  1. Gere um conjunto de m vetores aleatórios i.i.d, digamos $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$.
  2. Calcule $h \approx \frac{1}{n} \sum_{i=1}^{m} w(\mathbf{x}_i)$.

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, $\mathbf{X}_1, ..., \mathbf{X}_m$ de uma densidade candidata $g(.)$ com o mesmo suporte $f(.)$.
- Assim
  $h = \int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_k}^{b_k} w(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_k}^{b_k} w(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x}.$
- 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 vairiáveis 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