MAP2220 - Fundamentos de Análise Numérica

Prof. Pedro da Silva Peixoto
Prof. Antoine Laurain

EPrec 2º Sem. 2018

Método de Romberg aplicado ao cálculo de probabilidades

Data de entrega: 06/02/2019
(via email para ppeixoto@usp.br / laurain@ime.usp.br)

Introdução

Uma função densidade de probabilidade $f(x)$ descreve a distribuição de uma variável aleatória contínua x, em que a probabilidade de x estar entre um intervalo (a, b) é determinado pela integral

$$\mathbb{P}(a < x < b) = \int_a^b f(x)dx.$$

A teoria das probabilidades é de grande importância em aplicações no dia a dia, pois fornece confiabilidade na tomada de decisões. Exemplos práticos são o desenvolvimento de muitos produtos de consumo, tais como automóveis e eletro - eletrônicos, em que a teoria é utilizada para reduzir a probabilidade de falha, ou por exemplo, para analisar a probabilidade do efeito nos preços do petróleo influenciar drasticamente a economia, entre várias outras aplicações em economia, mercado, etc.

Objetivo

Desenvolver uma rotina baseada no método de Romberg para o cálculo da integral definida de uma dada função e aplicá-lo no cálculo de probabilidades.

Instruções

- O exercício-programa deverá ser feito individualmente e em linguagem C ou Python 3.x. EPs atrasados não serão aceitos.
- Entregar o código usado para as simulações computacionais (com extensão .c ou .py). Colocar o nome e número USP no arquivo, ex: JoãodaSilva123435.c

- Serão levados em conta na correção organização e comentários, portanto você deve usar funções no seu programa.
- Você também deve entregar um relatório explicando as análises que fez e os resultados obtibos (.pdf).

O Método de Romberg

O método de Romberg usa como base o método de n-trapézios, cujas propriedades foram analisadas em detalhe nas aulas da disciplina. O cálculo da integral pelo método dos trapézios é muito simples. Subdivide-se o intervalo de integração em n subintervalos de igual tamanho, aproximando-se a área definida pela integral da função em cada subintervalo pela área do trapézio obtido ao se interpolar os valores da função nos extremos do subintervalo por uma reta.

Vamos introduzir alguma notação. Seja $[a, b]$ o intervalo em que queremos integrar a função $f(x)$. Vamos definir os n subintervalos de $[a, b]$ de espaçamento $h = (b-a)/n$, através da introdução dos pontos $x_i = a + ih, i = 0, 1, ..., n$. Assim,

$$\int_a^b f(x)dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)dx.$$

Aproximando, $\int_{x_{i-1}}^{x_i} f(x)dx$ por $h\frac{f(x_{i-1})+f(x_i)}{2}$ em cada subintervalo obtemos a fórmula dos n-trapézios (com espaçamento $h = (b - a)/n$):

$$\int_a^b f(x)dx = T(h) = \sum_{i=1}^{n} h\frac{f(x_{i-1}) + f(x_i)}{2} = h\left(\frac{f(x_0)}{2} + \sum_{i=1}^{n-1} f(x_i) + \frac{f(x_n)}{2}\right).$$

O erro de aproximação na integral (para funções $f(x)$ em $C^2[a, b]$) é dado por:

$$E(h) = \int_a^b f(x)dx - T(h) = -(b - a)h^2 f''(y)/12 \quad ,$$

para algum y em $[a, b]$. Esta última relação mostra que o método de n-trapézios é convergente de ordem 2. Para funções de classe $C^{2m+2}[a, b]$ mostra-se através da fórmula de Euler-Maclaurin (veja por exemplo Stoer and Bulirsch, Introduction to Numerical Analysis) que:

$$E(h) = \int_a^b f(x)dx - T(h) = \sum_{k=1}^{m} c_k h^{2k} + \beta(h)h^{2m+2} \quad ,$$

em que os valores c_k são constantes que dependem da função $f(x)$, mas não do espaçamento h. Esta última expressão é uma "expansão assintótica do erro", e através dela podemos derivar esquemas convergentes de ordem mais alta.

Caso calculemos $T(h)$ e $T(h/2)$ (ou seja a fórmula dos trapézios para n e $2n$), podemos combinar estes valores de forma a eliminar o termo proporcional a h^2 na expansão do erro, obtendo:

$$S(h/2) = \frac{4T(h/2) - T(h)}{3} = \int_a^b f(x)dx - \bar{c}_2 h^4 - \bar{c}_3 h^6 - .... - \beta(h)h^{2m+2} \quad ,$$

tal que as novas constantes $\bar{c}_k$ resultam desta combinação. Este novo método, aproxima a integral de f pelo valor $S(h/2)$ com erro de ordem 4. Esta mesma ideia pode ser generalizada, combinando valores de $S(h/2)$ com $S(h/4)$ de forma a cancelar o erro de ordem 4 (formando a expressão $R(h/4) = \frac{16S(h/4) - S(h/2)}{15}$, verifique!), aumentando o ordem de convergência para 6, etc. O método de Romberg consiste em uma exploração recursiva desta ideia, sintetizada na tabela a seguir:

$$\begin{array}{llllll}
T(h) = T_{00} & & & & & \\
T(\frac{h}{2}) = T_{10} & T_{11} & & & & \\
T(\frac{h}{2^2}) = T_{20} & T_{21} & T_{22} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
T(\frac{h}{2^{n-1}}) = T_{n-1,0} & T_{n-1,1} & T_{n-1,2} & \dots & T_{n-1,n-1} & \\
T(\frac{h}{2^n}) = T_{n0} & T_{n1} & T_{n2} & \dots & T_{n,n-1} & T_{nn}
\end{array} \tag{1}$$

Cada coluna k do esquema acima é obtida da coluna anterior pela expressão

$$T_{ik} = \frac{4^k\, T_{i,k-1} - T_{i-1,k-1}}{4^k - 1} = T_{i,k-1} + \frac{T_{i,k-1} - T_{i-1,k-1}}{4^k - 1},$$

em que a segunda igualdade é conveniente do ponto de vista numérico por expressar o novo valor calculado como um valor da coluna anterior mais uma correção.

Todas as entradas da tabela representam uma aproximação para a integral, e o erro entre esta e T_{ik} decai proporcionalmente a $(h/2^{i-k})^{2k+2}$. Em particular, o erro entre T_{nn} e a integral decai como h^{2n+2}, o que nos dá um método de integração de ordem alta. Note que a coluna $k = 0$ é construída usando-se a fórmula dos trapézios e o espaçamento diminui por um fator de dois entre linhas consecutivas.

Implementação

O método de Romberg pode ser implementado iterativamente da seguinte forma. Fixe um valor de n (em geral n entre 4 e 6 é suficiente). Construa a tabela acima a partir dos valores $T(h_i)$, $i = 0, \ldots, n$, com $h_i = (b-a)/2^i$. Se o valor T_{nn} obtido for satisfatório (ver abaixo), pare. Senão, descarte h_0 e acrescente h_{n+1}, e repita o processo começando com $T(h_{i+1})$, $i = 0, \ldots, n$. E assim sucessivamente.

Critério de parada

Um critério de parada que funciona bem na prática consiste em especificar uma tolerância ϵ, e parar a execução do processo caso o erro relativo entre T_{nn} e $T_{n,n-1}$ seja menor do que ϵ. Isto significa que a última entrada da tabela não difere significativamente, dentro da precisão escolhida, de seu vizinho da coluna anterior, e estamos acreditando que daí em diante não haverá ganho de precisão. Um número máximo de iterações também deve ser especificado.

Função Romberg

Implemente o método de Romberg conforme descrito acima (em c). O subprograma para o cálculo da integral deve ser da forma

$$\mathbf{romb}(a, b, n, \epsilon, ITMAX)$$

em que os parâmetros são: a e b, extremos do intervalo de integração; n, número inteiro especificando quantos valores são usados na coluna $k = 0$ ($n+1$ valores); ϵ, tolerância; e $ITMAX$, número máximo de iterações. A execução deve parar quando pelo menos uma das condições abaixo for verificada:

- (i) $|T_{nn} - T_{n,n-1}| \leq \epsilon * |T_{nn}|$;
- (ii) número de iterações $= ITMAX$.

Trapézios

A fórmula dos trapézios pode ser implementada de forma eficiente em um processo iterativo, duplicando-se o número de intervalos a cada passo de modo que valores calculados anteriormente possam ser aproveitados. Se considerarmos $h_0 = (b-a)$, e $h_i = h_{i-1}/2$, $i = 1, 2, \ldots$ (i.e. $N_i = 2^i$), então

$$T(h_i) = \frac{1}{2}T(h_{i-1}) + h_i \sum_{j=1}^{2^{i-1}} f[a + (2j-1)h_i].$$

Note que para obter o próximo passo, precisamos calcular os valores de f apenas nos novos pontos acrescentados. Para o uso da fórmula dos trapézios, implemente uma rotina separada da forma

$$\textbf{trapz}(a, b, t, i)$$

em que o parâmetro de entrada e saída t recebe o valor calculado pela fórmula dos trapézios com 2^{i-1} intervalos e retorna o valor calculado pela fórmula dos trapézios com 2^i intervalos (no caso $i = 0$, a rotina simplesmente retorna em t o valor calculado pela fórmula dos trapézios com um intervalo). Esta rotina deve ser usada pela rotina do método de Romberg.

Distribuições

1. A **distribuição normal** é uma distribuição de probabilidade contínua com $x \in \mathbb{R}$, cuja função densidade de probabilidade é denotada por:

$$f(x; \mu; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2}$$

em que μ é a média, σ é o desvio padrão e σ^2 é a variância.

2. A função densidade de probabilidade de uma **distribuição chi** com suporte no intervalo $[0, \infty)$ é dada por

$$f(x;\, k) = \begin{cases} \dfrac{x^{(k/2-1)}e^{-x/2}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)}, & x > 0; \\ 0, & \text{caso contrário.} \end{cases} \tag{3}$$

em que $\Gamma(k/2)$ denota a função *Gamma* a qual tem forma fechada para k inteiro.

3. A função densidade de probabilidade de uma **distribuição exponencial** é dada por:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases} \tag{4}$$

Temos que $\lambda > 0$ é o parâmetro da distribuição com suporte em $x \in [0, \infty)$.

4. Se uma variável aleatória x tem uma **distribuição F** com parâmetros d_1 e d_2, escrevemos $x \sim F(d_1, d_2)$. Então, a função de densidade de probabilidade para x, com $x \in [0, \infty)$ e B é a função *beta*, é dada por

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1\, x)^{d_1}\ d_2^{d_2}}{(d_1\, x + d_2)^{d_1+d_2}}}}{x\, \mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}. \tag{5}$$

Problemas

1. Teste se seu método de Romberg está bem implementado para calcular as integrais:

$$\int_0^1 x^2 dx = \frac{1}{3} \quad e \quad \int_0^{0.995} \frac{dx}{1-x} = \ln(200)$$

Aplicações

2. As contas mensais de serviços públicos em determinada cidade são normalmente distribuídas, com média de $R\$100$ e desvio padrão de $R\$12$. Uma conta é escolhida aleatoriamente. Determine

   $a)$ A probabilidade de seu valor estar entre $R\$76$ e $R\$123$ ($\mathbb{P}(76 < x < 123)$).

   $b)$ A probabilidade do valor ser maior que $R\$120$ ($\mathbb{P}(x > 120)$).

3. O tempo de vida (em horas) de um transistor é uma variável aleatória T com distribuição exponencial. O tempo médio de vida do transistor é de 500 horas. Calcule:

   $a)$ A probabilidade de o transistor durar entre 200 e 500h ($\mathbb{P}(200 < T < 500)$).

   $b)$ A probabilidade de o transistor durar mais do que 500h ($\mathbb{P}(T > 500)$).

4. Uma amostra de 395 pessoas foi convidada a reportar o nível de educação mais alto que obtiveram. Resultados são mostrados na tabela abaixo:

	Ensino Médio	Graduação	Mestrado	Doutorado	Total
Mulheres	60	54	46	41	201
Homens	40	44	53	57	194
Total	100	98	99	98	395

Calculando a tabela com frequências esperadas, obtemos que $\mathcal{X}^2 = 8.0063$. Calcule:

$a)$ A probabilidade $\mathbb{P}(0 < x < 8.0)$.

$b)$ A probabilidade $\mathbb{P}(x > 8.0)$.

5. Ao realizar um teste estatístico de análise de variâncias (ANOVA), foi obtida uma estatística de teste F com valor 2,57. Assumindo $d_1 = 3$ e $d_2 = 246$ graus de liberdade, calcule a probabilidade $\mathbb{P}(x > 2.57)$.

Entrada

O seu programa deve ser geral, no sentido de prever a possibilidade de calcular probablidades para qualquer intervalo (inclusive quando a ou b forem mais ou menos infinito) para todas as distribuições descritas acima. Os seus parâmetros para o método Romberg podem estar definidos internamente no código, sendo apenas necessário fornecer um arquivo de entrada (ou entrada via linha de comando) contendo:

- o número da distribuição escolhida (1 a 4) ou função a ser integrada (Problema 1, integral 1 ou 2),
- os valores a,b para o intervalo de integração, onde deve ser possível usar como entrada "inf" e "-inf" para designar infinito.
- os parâmetros da distribuição.

Saída

Para esse EP, não é obrigatório a entrega de um relatório junto com o código, então a saída do seu programa deve ser bem clara, e conter uma impressão de um cabeçalho com seu nome e número usp, e para cada problema proposto, a impressão da probabilidade calculada.

Dicas

Podem (devem) ser usadas as propriedades das distribuições para transformar o problema de integração para um intervalo compacto (caso o intervalo inicial não o seja).

Bons Estudos!

Referências

[1] J. Stoer, R. Bulirsch, *Introduction to numerical analysis,* Vol 12, Springer Science & Business Media, 2013.