CE-718 (MCI)
Contexto e idéias iniciais

PJ

LEG: Laboratório de Estatística e Geoinformação

Out, 2009

Principais usos em estatística

- avaliação de funções de verossimilhança
  - maximização
  - avaliação para obter *posterioris*
  - avaliação em simulações
- cálculo de esperanças
  - estimadores expressos como esperanças
  - predição de efeitos
  - predição (temporal/espacial)
  - predição de efeitos aleatórios
  - imputação
  - ...

Inferência via verossimilhança

- Soluções analíticas
- Soluções numéricas
  - convencionais / "bem comportadas"
  - "trabalhosas"
  - impossíveis!

Inferência via verossimilhança

- Soluções analíticas
- Soluções numéricas
  - convencionais / ”bem comportadas”
  - ”trabalhosas”
  - impossíveis!

Soluções numéricas

Tópicos em algorítmos numéricos

- implementações disponíveis
- programação (eficiente) da função objetivo (ex.)
- escolha do algorítmo
- ”calibragem”do algorítmo

Soluções numéricas

Tópicos em algorítmos numéricos

- implementações disponíveis
- programação (eficiente) da função objetivo (ex.)
- escolha do algorítmo
- ”calibragem”do algorítmo

Soluções numéricas

Tópicos em algorítmos numéricos

- implementações disponíveis
- programação (eficiente) da função objetivo (ex.)
- escolha do algorítmo
- ”calibragem”do algorítmo

Soluções numéricas

Tópicos em algorítmos numéricos

- implementações disponíveis
- programação (eficiente) da função objetivo (ex.)
- escolha do algorítmo
- ”calibragem”do algorítmo

Soluções numéricas (cont.)

Cálculos numéricos e derivadas

- diferenciação
- diferenciação automática
- gradientes
- hessianos

Soluções numéricas (cont.)

Cálculos numéricos e derivadas

- diferenciação
- diferenciação automática
- gradientes
- hessianos

Soluções numéricas (cont.)

Cálculos numéricos e derivadas

- diferenciação
- diferenciação automática
- gradientes
- hessianos

Soluções numéricas (cont.)

Cálculos numéricos e derivadas

- diferenciação
- diferenciação automática
- gradientes
- hessianos

Abordagens para integração

Abordagens determinísticas

- Soluções analíticas
- Equações diferenciais

$$y = \int_a^b f(x)\,dx$$

é solução de:

$$\frac{\mathrm{dy}}{\mathrm{dx}} = f(x), \quad y(a) = 0$$

- Regras de quadratura

$$\int_a^b f(x)\,dx \approx \frac{b-a}{N}\sum_{i=1}^{N} f(a + (i-0.5)(b-a)/N)$$

Abordagens para integração

Abordagens determinísticas

- Soluções analíticas
- Equações diferenciais

$$y = \int_a^b f(x)\, dx$$

é solução de:

$$\frac{\mathrm{dy}}{\mathrm{dx}} = f(x), \quad y(a) = 0$$

- Regras de quadratura

$$\int_a^b f(x)\, dx \approx \frac{b-a}{N} \sum_{i=1}^{N} f(a + (i - 0.5)(b-a)/N)$$

Abordagens para integração

Abordagens determinísticas

- Soluções analíticas
- Equações diferenciais

$$y = \int_a^b f(x)\, dx$$

é solução de:

$$\frac{\mathrm{dy}}{\mathrm{dx}} = f(x), \quad y(a) = 0$$

- Regras de quadratura

$$\int_a^b f(x)\, dx \approx \frac{b-a}{N} \sum_{i=1}^{N} f(a + (i - 0.5)(b - a)/N)$$

Aproximação de integrais – quadratura

$$\int_a^b f(x)\,dx = \frac{b-a}{N}\sum_{i=1}^{N} f(a + (i-0.5)(b-a)/N) + O(h^2)$$

- **lei do ponto médio:**
  aproximação por *função escada*
- **lei de Simpson:**
  aproximação de maior ordem
- **Quadratura Gaussiana:**
  acurácia de maior ordem através de escolha criteriosa de pontos onde avaliar a função

$$\int_a^b f(x)\,dx = \int_a^b g(x)w(x)\,dx = \sum_{i=1} nw_i g(x_i)$$

Aproximação de integrais – quadratura

$$\int_a^b f(x)\,dx = \frac{b-a}{N}\sum_{i=1}^{N} f(a+(i-0.5)(b-a)/N) + O(h^2)$$

- **lei do ponto médio:**
  aproximação por *função escada*
- **lei de Simpson:**
  aproximação de maior ordem
- **Quadratura Gaussiana:**
  acurácia de maior ordem através de escolha criteriosa de pontos onde avaliar a função

$$\int_a^b f(x)\,dx = \int_a^b g(x)w(x)\,dx = \sum_{i=1} nw_i g(x_i)$$

Aproximação de integrais – quadratura

$$\int_a^b f(x)\,dx = \frac{b-a}{N}\sum_{i=1}^{N} f(a+(i-0.5)(b-a)/N) + O(h^2)$$

- **lei do ponto médio:**
  aproximação por *função escada*
- **lei de Simpson:**
  aproximação de maior ordem
- **Quadratura Gaussiana:**
  acurácia de maior ordem através de escolha criteriosa de pontos onde avaliar a função

$$\int_a^b f(x)\,dx = \int_a^b g(x)w(x)\,dx = \sum_{i=1} nw_i g(x_i)$$

Comentários

- ganhos expressivos para QG em termos de avaliação de função e acurácia ...
- ... mas especialmente e notadamente em funções suaves!
- viável para 1 ou 2 dimensões (talvez até +/- 4 em problemas "bem comportados")
- para integrais de alta dimensão:
  - número de avaliações da função pode se tornar proibitivo
  - aproximação da função por uma função integrável
  - abordagem estatística $\longrightarrow$ estimador da integral

  - viés e dimensão
  - variância
  - viés *vs* variância (ou ou outro!!??)

Comentários

- ganhos expressivos para QG em termos de avaliação de função e acurácia . . .
- . . . mas especialmente e notadamente em funções suaves!
- viável para 1 ou 2 dimensões (talvez até +/- 4 em problemas "bem comportados")
- para integrais de alta dimensão:
  - número de avaliações da função pode se tornar proibitivo
  - aproximação da função por uma função integrável
  - abordagem estatística $\longrightarrow$ estimador da integral

  - viés e dimensão
  - variância
  - viés *vs* variância (ou ou outro!!??)

Comentários

- ganhos expressivos para QG em termos de avaliação de função e acurácia ...
- ...mas especialmente e notadamente em funções suaves!
- viável para 1 ou 2 dimensões (talvez até +/- 4 em problemas "bem comportados")
- para integrais de alta dimensão:
  - número de avaliações da função pode se tornar proibitivo
  - aproximação da função por uma função integrável
  - abordagem estatística $\longrightarrow$ estimador da integral
  - viés e dimensão
  - variância
  - viés *vs* variância (ou ou outro!!??)

Comentários

- ganhos expressivos para QG em termos de avaliação de função e acurácia ...
- ...mas especialmente e notadamente em funções suaves!
- viável para 1 ou 2 dimensões (talvez até +/- 4 em problemas ”bem comportados”)
- para integrais de alta dimensão:
  - número de avaliações da função pode se tornar proibitivo
  - aproximação da função por uma função integrável
  - abordagem estatística $\longrightarrow$ estimador da integral

  - viés e dimensão
  - variância
  - viés *vs* variância (ou ou outro!!??)

Quadratura Gaussiana

- Pesos e pontos : diferentes critérios, tabelas e algorítmos para atribuição
  - quadraturas definidas no intervalo (-1, 1)

$$\int_a^b g(t)dt = \frac{b-a}{2}\int_{-1}^{1} g(\frac{a+b+(b-a)x}{2}dx = \int_{-1}^{1} f(x)dx$$

$$t = a + (b-a)(x+1)/2$$

$$x = -1 + 2(t-a)/(b-a)$$

$dt = (b-a)/2dx$

  - implementação em `statmod:gauss.quad`
  - quadratura com pontos baseados em uma dist. normal:

$$\int f(x)\phi(x)dx \approx \sum_{i=1}^{N} w_i f(x_i)$$

    - $\phi(x)$ é a densidade de uma normal (multivariada) padronizada
- implementação em `statmod:gauss.quad.prob`

Quadratura Gaussiana

- Pesos e pontos : diferentes critérios, tabelas e algorítmos para atribuição
  - quadraturas definidas no intervalo (-1, 1)

$$\int_a^b g(t)dt = \frac{b-a}{2}\int_{-1}^{1} g(\frac{a+b+(b-a)x}{2}dx = \int_{-1}^{1} f(x)dx$$

$$t = a + (b-a)(x+1)/2$$

$$x = -1 + 2(t-a)/(b-a)$$

$$dt = (b-a)/2dx$$

  - implementação em `statmod:gauss.quad`
  - quadratura com pontos baseados em uma dist. normal:

$$\int f(x)\phi(x)dx \approx \sum_{i=1}^{N} w_i f(x_i)$$

    - $\phi(x)$ é a densidade de uma normal (multivariada) padronizada
  - implementação em `statmod:gauss.quad.prob`

Quadratura Gaussiana

- Pesos e pontos : diferentes critérios, tabelas e algorítmos para atribuição
  - quadraturas definidas no intervalo (-1, 1)

$$\int_a^b g(t)dt = \frac{b-a}{2}\int_{-1}^{1} g(\frac{a+b+(b-a)x}{2}dx = \int_{-1}^{1} f(x)dx$$

$$t = a + (b-a)(x+1)/2$$

$$x = -1 + 2(t-a)/(b-a)$$

$$dt = (b-a)/2dx$$

  - implementação em `statmod:gauss.quad`
  - quadratura com pontos baseados em uma dist. normal:

$$\int f(x)\phi(x)dx \approx \sum_{i=1}^{N} w_i f(x_i)$$

    - $\phi(x)$ é a densidade de uma normal (multivariada) padronizada
  - implementação em `statmod:gauss.quad.prob`

Quadratura Gaussiana

- Pesos e pontos : diferentes critérios, tabelas e algorítmos para atribuição
  - quadraturas definidas no intervalo (-1, 1)

$$\int_a^b g(t)dt = \frac{b-a}{2}\int_{-1}^{1} g(\frac{a+b+(b-a)x}{2}dx = \int_{-1}^{1} f(x)dx$$

$$t = a + (b-a)(x+1)/2$$

$$x = -1 + 2(t-a)/(b-a)$$

$$dt = (b-a)/2dx$$

  - implementação em `statmod:gauss.quad`
  - quadratura com pontos baseados em uma dist. normal:

$$\int f(x)\phi(x)dx \approx \sum_{i=1}^{N} w_i f(x_i)$$

    - $\phi(x)$ é a densidade de uma normal (multivariada) padronizada
  - implementação em `statmod:gauss.quad.prob`

Integrais multidimensionais

Integrando $f(x, z)$ em uma região quadrada

- Uso recursivo de quadratura:
  considere mesmos nós e pesos nas duas dimensões:

$$\int f(x_j, z)\, dz \approx \sum_i w_i f(x_j, z_i)$$

  e desta forma:

$$\int \int f(x,z) dz dx \approx \sum_j w_j \sum_i w_i f(x_j, z_i) = \sum_j \sum_i f(x_j, z_i)$$

- e pode-se considerar o argumento recursivamente para mais dimensões.
- Para manter acurácia: N^d avaliações de função.
  Exs: faça algumas contas!!

Integrais multidimensionais

Integrando $f(x, z)$ em uma região quadrada

- Uso recursivo de quadratura:
  considere mesmos nós e pesos nas duas dimensões:

$$\int f(x_j, z)\, dz \approx \sum_i w_i f(x_j, z_i)$$

  e desta forma:

$$\int \int f(x, z) dz dx \approx \sum_j w_j \sum_i w_i f(x_j, z_i) = \sum_j \sum_i f(x_j, z_i)$$

- e pode-se considerar o argumento recursivamente para mais dimensões.
- Para manter acurácia: N^d avaliações de função.
  Exs: faça algumas contas!!

Integrais multidimensionais

Integrando $f(x,z)$ em uma região quadrada

- Uso recursivo de quadratura:
  considere mesmos nós e pesos nas duas dimensões:

$$\int f(x_j,z)\,dz \approx \sum_i w_i f(x_j,z_i)$$

  e desta forma:

$$\int\int f(x,z)dzdx \approx \sum_j w_j \sum_i w_i f(x_j,z_i) = \sum_j \sum_i f(x_j,z_i)$$

- e pode-se considerar o argumento recursivamente para mais dimensões.
- Para manter acurácia: N^d avaliações de função.
  Exs: faça algumas contas!!

Outra abordagem: Aproximando o integrando

Princípio: Aproximar $f(x)$ por uma função tratável ("facilmente"integrável).

- Diversas abordagens possíveis
- Destaca-se a *Aproximação de Laplace*

Aplicação típica:

- considere $f_y(y) = \int f(y, b)\ d(b)$
- argumento baseado em Taylor
- $f_y(y) \approx f(y, \hat{b}_y) \frac{(2\pi)^{d/2}}{|H|^{1/2}}$

Elementos requeridos:
H (hessiano de $-\log(f)$ em relação a b avaliado em $y, \hat{b}_y$) e $\hat{b}_y$ (método de Newton)

Outra abordagem: Aproximando o integrando

Princípio: Aproximar $f(x)$ por uma função tratável ("facilmente"integrável).

- Diversas abordagens possíveis
- Destaca-se a *Aproximação de Laplace*

Aplicação típica:

- considere $f_y(y) = \int f(y,b)\ d(b)$
- argumento baseado em Taylor
- $f_y(y) \approx f(y, \hat{b}_y) \frac{(2\pi)^{d/2}}{|H|^{1/2}}$

Elementos requeridos:
H (hessiano de $-\log(f)$ em relação a b avaliado em $y, \hat{b}_y$) e $\hat{b}_y$ (método de Newton)

Integração de Monte Carlo

- Considere integrar $f(x)$ em uma região R de volume $V(R)$.
- A integral pode ser escrita como uma esperança:

$$I_R = \int_R f(x)\, dx = \mathbb{E}\{f(X)\}V(R)$$

- $X \sim U(R)$.
- Estimador: gerando-se N vetores aleatórios uniformes x_i:

$$\hat{I}_R = \frac{V(R)}{N} \sum_{i=1}^{N} f(x_i)$$

- $\hat{I}_R$ é não viciado com

$$\text{Var}(\hat{I}_R) = \frac{V^2(R)}{N^2}\, N\ \text{Var}\{f(x)\}$$

Integração de Monte Carlo

- Considere integrar $f(x)$ em uma região R de volume $V(R)$.
- A integral pode ser escrita como uma esperança:

$$I_R = \int_R f(x)\, dx = \mathbb{E}\{f(X)\} V(R)$$

- $X \sim U(R)$.
- Estimador: gerando-se N vetores aleatórios uniformes x_i:

$$\hat{I}_R = \frac{V(R)}{N} \sum_{i=1}^{N} f(x_i)$$

- $\hat{I}_R$ é não viciado com

$$\text{Var}(\hat{I}_R) = \frac{V^2(R)}{N^2} N \text{ Var}\{f(x)\}$$

Integração de Monte Carlo

- Considere integrar $f(x)$ em uma região R de volume $V(R)$.
- A integral pode ser escrita como uma esperança:

$$I_R = \int_R f(x)\, dx = \mathbb{E}\{f(X)\}V(R)$$

- $X \sim U(R)$.
- Estimador: gerando-se N vetores aleatórios uniformes x_i:

$$\hat{I}_R = \frac{V(R)}{N} \sum_{i=1}^{N} f(x_i)$$

- $\hat{I}_R$ é não viciado com

$$\mathrm{Var}(\hat{I}_R) = \frac{V^2(R)}{N^2}\, N\ \mathrm{Var}\{f(x)\}$$

Comentários

- para $d = 4$ convergência estocástica supera a determinística de aproximação de ponto médio;
- para funções suaves QG permanece superior mesmo para dimensões maiores;
- porém em QG viés depende de d enquanto que em IMC é independente de dimensão.

Comentários

- para $d = 4$ convergência estocástica supera a determinística de aproximação de ponto médio;
- para funções suaves QG permanece superior mesmo para dimensões maiores;
- porém em QG viés depende de d enquanto que em IMC é independente de dimensão.

Comentários

- para $d = 4$ convergência estocástica supera a determinística de aproximação de ponto médio;
- para funções suaves QG permanece superior mesmo para dimensões maiores;
- porém em QG viés depende de d enquanto que em IMC é independente de dimensão.

Variações

Aprimoramentos (variância, etc)

- Integração de Monte Carlo Estratificada, Quasi-Monte Carlo, …
- Variante: integrando fatora-se num produto de um termo $\phi(x)$ e uma f.d.p. $g(x)$

$$\int \phi(x)g(x)\,dx = \mathbb{E}_f\{\phi(x)\} \approx \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$

- Problemas com a proporção de valores de $\phi(x)$ com nenhuma ou pouca contribuição para integral
- Uma alternativa: amostragem por importância (*importance sampling*)

$$\int \phi(x)g(x)\,dx \approx \hat{\hat{I}} = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\frac{g(x_i)}{h(x_i)}$$

$h(x)$ tem mesmo suporte que $g(x)$ mas que concentra pontos em regiões relevantes.

Variações

Aprimoramentos (variância, etc)

- Integração de Monte Carlo Estratificada, Quasi-Monte Carlo, ...
- Variante: integrando fatora-se num produto de um termo $\phi(x)$ e uma f.d.p. $g(x)$

$$\int \phi(x)g(x)\,dx = \mathbb{E}_f\{\phi(x)\} \approx \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$

- Problemas com a proporção de valores de $\phi(x)$ com nenhuma ou pouca contribuição para integral
- Uma alternativa: amostragem por importância (*importance sampling*)

$$\int \phi(x)g(x)\,dx \approx \hat{\hat{I}} = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\frac{g(x_i)}{h(x_i)}$$

$h(x)$ tem mesmo suporte que $g(x)$ mas que concentra pontos em regiões relevantes.

Variações

Aprimoramentos (variância, etc)

- Integração de Monte Carlo Estratificada, Quasi-Monte Carlo, ...
- Variante: integrando fatora-se num produto de um termo $\phi(x)$ e uma f.d.p. $g(x)$

$$\int \phi(x)g(x)\, dx = \mathbb{E}_f\{\phi(x)\} \approx \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$

- Problemas com a proporção de valores de $\phi(x)$ com nenhuma ou pouca contribuição para integral
- Uma alternativa: amostragem por importância (*importance sampling*)

$$\int \phi(x)g(x)\, dx \approx \hat{\hat{I}} = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\frac{g(x_i)}{h(x_i)}$$

$h(x)$ tem mesmo suporte que $g(x)$ mas que concentra pontos em regiões relevantes.

Variações

Aprimoramentos (variância, etc)

- Integração de Monte Carlo Estratificada, Quasi-Monte Carlo, ...
- Variante: integrando fatora-se num produto de um termo $\phi(x)$ e uma f.d.p. $g(x)$

$$\int \phi(x)g(x)\, dx = \mathbb{E}_f\{\phi(x)\} \approx \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)$$

- Problemas com a proporção de valores de $\phi(x)$ com nenhuma ou pouca contribuição para integral
- Uma alternativa: amostragem por importância (*importance sampling*)

$$\int \phi(x)g(x)\, dx \approx \hat{\hat{I}} = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i)\frac{g(x_i)}{h(x_i)}$$

$h(x)$ tem mesmo suporte que $g(x)$ mas que concentra pontos em regiões relevantes.

Laplace e Amostragem por importância

- Como escolher $h(\cdot)$?
- seguindo Laplace, aproximar por uma NMV
- aproximação por NMV com facilidades para simular

$$x = \hat{x} + R^{-1}z \qquad\qquad R'R = H$$

- re-expressando em função de z

$$\int \phi(x)g(x)\,dx \approx \frac{(2\pi)^{d/2}}{n|R|}\sum_{i=1}^{n}\phi(\hat{x}+R^{-1}z_i)f(\hat{x}+R^{-1}z_i)\exp\{z_i'z_i/2\}$$

Detalhes adicionais relevantes na aplicação

Laplace e Amostragem por importância

- Como escolher $h(\cdot)$?
- seguindo Laplace, aproximar por uma NMV
- aproximação por NMV com facilidades para simular

$$x = \hat{x} + R^{-1} z \qquad R'R = H$$

- re-expressando em função de z

$$\int \phi(x) g(x)\, dx \approx \frac{(2\pi)^{d/2}}{n|R|} \sum_{i=1}^{n} \phi(\hat{x}+R^{-1}z_i) f(\hat{x}+R^{-1}z_i) \exp\{z_i' z_i/2\}$$

Detalhes adicionais relevantes na aplicação

Laplace e Amostragem por importância

- Como escolher $h(\cdot)$?
- seguindo Laplace, aproximar por uma NMV
- aproximação por NMV com facilidades para simular

$$x = \hat{x} + R^{-1}z \qquad\qquad R'R = H$$

- re-expressando em função de z

$$\int \phi(x)g(x)\,dx \approx \frac{(2\pi)^{d/2}}{n|R|}\sum_{i=1}^{n}\phi(\hat{x}+R^{-1}z_i)f(\hat{x}+R^{-1}z_i)\exp\{z_i'z_i/2\}$$

Detalhes adicionais relevantes na aplicação

Exemplo 1

Considere o modelo de regressão de Poisson com intercepto aleatório

$$Y_i|b \sim Poi(\mu_i)$$
$$\log(\mu_i) = \beta_0 + \beta_1 x_i + b_i$$
$$b_i \sim N(0, \sigma_b^2)$$

- verossimilhança
- aproximação por métodos de quadratura
- aproximação de Laplace
- Integração de Monte Carlo
- Amostragem por importância (Laplace)