Método de Monte Carlo para propagação de incertezas com covariância

Alexandre Suaide

Imagine que se quer calcular uma grandeza *y* que é resultado de uma função de diversos parâmetros *x*. Porém esses parâmetros *x* possuem incertezas e covariâncias entre si. Como calcular a incerteza em *y* pelo método de Monte Carlo considerando a covariância entre os parâmetros?

O Método de Monte Carlo para propagar incertezas consiste em sortear o conjunto de parâmetros *x* segundo suas F.D.P. Isso é fácil de fazer se todos os elementos forem independentes. Neste caso, basta sortear cada parâmetro segundo sua F.D.P. e, para este conjunto de parâmetros sorteados, calcula-se *y*. Repete-se isso um número elevado de vezes de modo a obter a F.D.P de *y*, de onde estima-se a sua incerteza. No caso de haver covariância entre os parâmetros, o sorteio do conjunto de parâmetros *x* é feito através de uma operação matricial, tal que:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1^0 \\ x_2^0 \\ \vdots \\ x_n^0 \end{pmatrix} + L \begin{pmatrix} gaus(0,1) \\ gaus(0,1) \\ \vdots \\ gaus(0,1) \end{pmatrix}$$

onde gaus(0,1) corresponde a um número aleatório com distribuição gaussiana de média 0 e desvio padrão 1. Isso tem implícito que as variáveis *x* possuem F.D.P. gaussianas. Se não possuírem, basta substituir o sorteio gaussiano pela distribuição apropriada, desde que tenha média 0 e desvio padrão 1. Os elementos $x_1^0, x_2^0, \dots, x_n^0$ correspondem aos valores médios de *x*. Os elementos $x_1, x_2, \dots, x_n$, correspondem ao conjunto sorteado para uma dada iteração. Calcula-se *y* com este conjunto de parâmetros. Repete-se a operação muitas vezes para encontrar a F.D.P. de *y*.

A matriz *L* é a matriz de Cholesky [1] da matriz de covariância. A matriz de covariância é dada por:

$$A = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}_{12} & \cdots & \mathrm{cov}_{1n} \\ \mathrm{cov}_{21} & \sigma_2^2 & & \\ \vdots & & \ddots & \\ \mathrm{cov}_{n1} & & & \sigma_n^2 \end{pmatrix}$$

Como a matriz de covariância é simétrica, ela pode ser decomposta de tal forma que [1]:

$$A = LL^T$$

Onde L é chamada matriz de Cholesky. L é uma matriz triangular. Há vários algoritmos para computar a matrix de Cholesky. Como a matriz de covariância é composta de números reais, utilizando-se o método de Cholesky-Crout [1] pode-se escrever que os elementos de L são dados por:

$$L_{jj} = \sqrt{A_{jj} - \sum_{k=1}^{j-1} L_{jk}^2}$$

$$L_{ij} = \frac{1}{L_{jj}}\left(A_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk}\right)$$

Este cálculo é iterativo. Começa-se pelo elemento $L_{11} = \sqrt{A_{11}} = \sigma_1$ e segue-se. Por exemplo, o elemento (2,1) da matriz é dado por:

$$L_{21} = \frac{1}{L_{11}}\left(A_{12} - L_{21}\underbrace{L_{12}}_{0}\right) = \frac{A_{12}}{L_{11}} = \frac{\mathrm{cov}_{21}}{\sigma_1}$$

E assim por diante. Segue abaixo um algoritmo em C para o cálculo da matriz de Cholesky.

```
CalcCholesky(int NP, double* fCov, double* fCovSqrt)
{
  double *C = fCovSqrt;
  double *V = fCov;
  // calculate sqrt(V) as lower diagonal matrix
  for( int i = 0; i < NP; ++i )
  {
    for( int j = 0; j < NP; ++j )
    {
      C[i*NP+j] = 0;
    }
  }

  for( int j = 0; j < NP; ++j )
  {
    // diagonal terms first
    double Ck = 0;
    for( int k = 0; k < j; ++k )
    {
      Ck += C[j*NP+k] * C[j*NP+k];
    } // k
    C[j*NP+j] = sqrt( fabs( V[j*NP+j] - Ck ) );

    // off-diagonal terms
    for( int i = j+1; i < NP; ++i )
    {
      Ck = 0;
      for( int k = 0; k < j; ++k )
      {
        Ck += C[i*NP+k] * C[j*NP+k];
      } //k
      if(C[j*NP+j]!=0 ) C[i*NP+j] = ( V[i*NP+j] - Ck ) / C[j*NP+j];
      else C[i*NP+j] = 0;
    }// i
  } // j
}
```

onde fCov é um vetor unidimensional de NP*NP elementos onde está a matriz de covariância e fCovSqrt é um vetor unidimensional de NP*NP elementos com a matriz de Cholesky resultante.

Uma vez de posse da matriz de Cholesky, basta sortear o conjunto de parâmetros de acordo com a expressão matricial no começo deste texto e, com este conjunto de parâmetros, calcular a função *y*. O conjunto de parâmetros sorteado vai conter os efeitos de todas as covariância e incertezas das variáveis *x*.

Referências

[1] http://en.wikipedia.org/wiki/Cholesky_decomposition