

## データ解析 課題 第3回

概要： MCMC 法による乱数生成

提出方法： 結果をレポート（紙）にまとめて西 8 W のレポートボックス「データ解析」へ提出．

締め切り： 6月2日（月）の 15:00

### 3.1 逆関数法の導出

区間  $(0, 1)$  の一様分布に従う独立な確率変数列  $U_1, U_2, \dots \sim U(0, 1)$  が利用できると仮定する．ある指定した分布  $F(x)$  (密度関数  $f(x)$ ) に従う  $X_1, X_2, \dots$  をどうやって得るかを考える． $X = F^{-1}(U)$  とおけば， $X$  の分布関数が  $F(x)$  であることを示せ．

### 3.2 逆関数法の数値例

区間  $(-1, 1)$  の確率変数  $X$  の密度関数が  $f(x) \propto x + 1$ ,  $-1 \leq x \leq 1$  で与えられるとする．このとき，下記の計算をおこなう R のスクリプトを作成して，実行せよ．スクリプトの定義と実行結果がわかるように「コンソール出力」を示すこと．

- 逆関数法を用いて  $f(x)$  に従う独立な疑似乱数を 10000 個生成せよ。
- そのヒストグラムを図示せよ。
- 生成した  $X$  の平均と分散を計算し、それらの理論値と比較せよ。

### 3.3 棄却法の導出

$f(x)$ ,  $g(x)$  を連続分布の密度関数とする。  $U \sim U(0, 1)$  と  $V \sim g(v)$  にしたがう乱数を利用して、  $X \sim f(x)$  を生成したい。次のアルゴリズムでこれが実現できることを示せ。

1. すべての  $x$  で  $f(x) \leq cg(x)$  となるような  $c$  を選ぶ。
2.  $U \sim U(0, 1)$  と  $V \sim g(v)$  をそれぞれ 1 個生成。
3.  $U \leq \frac{f(V)}{cg(V)}$  なら  $X = V$  として終了。そうでなければ 2へ戻る。

### 3.4 棄却法の数値例

区間  $(-1, 1)$  の確率変数  $X$  の密度関数が  $f(x) \propto \sqrt{1-x^2}$ ,  $-1 \leq x \leq 1$  で与えられるとする。このとき、下記の計算をおこなう R のスクリプトを作成して、実行せよ。スクリプトの定義と実行結果がわかるように「コンソール出力」を示すこと。

- 棄却法を用いて  $f(x)$  に従う独立な疑似乱数を 10000 個生成せよ。ただし，棄却法で用いる  $g(x)$  と  $c$  については，各自適切なものを検討し，その妥当性についてレポートに記述する。
- そのヒストグラムを図示せよ。
- 生成した  $X$  の平均と分散を計算し，それらの理論値と比較せよ。

### 3.5 MCMC 法（メトロポリス・ヘイスティングス アルゴリズム）の導出

系列  $X_1, X_2, \dots$  を次のアルゴリズムで生成する。このマルコフ連鎖の推移確率を決める条件付確率分布  $p(x_{t+1}|x_t)$  を求め，それが詳細釣り合  $p(x_{t+1}|x_t)f(x_t) = p(x_t|x_{t+1})f(x_{t+1})$  を満たすことを示せ。

1. 条件付密度関数  $q(v|x)$  を決める。初期値  $X_1$  を選び  $t = 1$  とする。
2.  $U \sim U(0, 1)$  と  $V \sim q(v|X_t)$  をそれぞれ 1 個生成。
3. 次の数を計算。

$$\alpha(V, X_t) = \min \left\{ 1, \frac{q(X_t|V)f(V)}{q(V|X_t)f(X_t)} \right\}$$

4.  $U \leq \alpha(V, X_t)$  なら  $X_{t+1} = V$  とする。そうでないなら  $X_{t+1} = X_t$  とする。
5.  $t$  の値をひとつ増やして 2 へもどる。

## 3.6 MCMC 法の数値例

確率変数  $X$  は平均 0 , 分散 1 の正規分布に従う確率変数とする . すなわち , 密度関数は ,  $f(x) \propto \exp(-x^2/2)$  で与えられる . 下記の計算をおこなう R のスクリプトを作成して , 実行せよ . スクリプトの定義と実行結果がわかるように「コンソール出力」を示すこと .

- MCMC 法を用いて  $f(x)$  に従う系列  $X_1, X_2, \dots$  を生成せよ . 系列長は 10000 とする . 提案分布  $q(v|x)$  は各自適切なものを検討し , その妥当性についてレポートに記述する . ただし ,  $q(v|x) \propto \exp(-v^2/2)$  などのように ,  $q(v|x)$  が  $x$  に依存しないものは認めない .
- 生成した系列  $x_1, x_2, \dots, x_{10000}$  の折れ線プロット ( 横軸  $t$ , 縦軸  $x_t$  ) を図示せよ . MCMC の移動頻度を表示せよ . 系列が十分に定常といえるかを議論せよ .
- 生成した系列のヒストグラムを図示せよ .
- 生成した  $X$  の平均と分散を計算して , それらの理論値と比較せよ .