

目標は正規分布に従うサンプルを得る事である.

確率変数 Y が正規分布に従うとは

$$P(Y < y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

なる事である. 右辺の積分は書き下せないで、むしろ密度関数

$$\frac{d}{dy} P(Y < y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

で与えられていると見なせる.

1 失敗と光明

$[0, 1]$ 区間上の一様分布に従う確率変数 X のサンプルは例えば

```
double urand(){ return (double)rand()/RANDMAX; }
```

によって容易に得られるので、まずは X に適当な関数 f を合成する事によって $Y = f(X)$ として正規分布を実現できないかを考える.

取り敢えず計算を試みれば

$$\begin{aligned} \frac{d}{dy} P(f(X) < y) &= \lim_{h \rightarrow 0} \frac{1}{h} P(y \leq f(X) < y + h) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} P(f^{-1}(y) \leq X < f^{-1}(y + h)) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} (f^{-1}(y + h) - f^{-1}(y)) \\ &= \frac{d}{dy} f^{-1}(y) \end{aligned}$$

であるので、密度関数 g に従う確率変数を $f(X)$ の形で実現するためには f は g の原始関数の逆関数である必要がある. 特に正規分布の密度関数 $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ の原始関数は初等的でないで正規分布を直接この方法で実現するのは現実的でない.

この方法で実現できる為にはもっと単純な密度関数の分布で無ければならない. 例えば正規分布の密度関数より e の肩の次数が 1 低い $g(x) = \frac{1}{\lambda} e^{-x/\lambda}$ を密度関数として持つ指数分布はこの方法で実現可能である. 実際、原始関数が $G(x) = 1 - e^{-x/\lambda}$ ととれるので $f(x) = G^{-1}(x) = -\lambda \log(1 - x)$. よって $f(X) = -\lambda \log(1 - X)$ とすればよい. 特に X と $1 - X$ は分布が等しいので $\tilde{f}(X) = -\lambda \log X$ とする事によって容易に実現できる

```
double lambdaexp(){ return -lambda * log( urand() ); }
```

じつはこの指数分布が正規分布と関係があり、この指数分布の実現が正規分布の実現への足がかりとなる.

2 正規分布と指数分布

Y_1, Y_2 を正規分布に従う互いに独立な確率変数, Z, Θ も互いに独立で Z が $\lambda = 2$ の指数分布, Θ は $[0, 2\pi]$ 上の一様分布に従う確率変数とすれば (Y_1, Y_2) の分布と $(\sqrt{Z} \cos \Theta, \sqrt{Z} \sin \Theta)$ の分布は等しい.

Proof. 局座標変換により

$$\begin{aligned} E[f(X, Y)] &= \int \int f(x, y) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \frac{1}{2\pi} \int \int f(x, y) e^{-(x^2+y^2)/2} dx dy \\ &= \frac{1}{2\pi} \int \int f(r \cos \theta, r \sin \theta) e^{-(r^2)/2} r dr d\theta \end{aligned}$$

$r^2 = z$ で変数変換すれば

$$\begin{aligned} E[f(X, Y)] &= \int \int f(\sqrt{z} \cos \theta, \sqrt{z} \sin \theta) \frac{1}{2} e^{-z/2} dz \frac{1}{2\pi} d\theta \\ &= E[f(\sqrt{Z} \cos \Theta, \sqrt{Z} \sin \Theta)] \end{aligned}$$

□

3 ボックス=ミュラー法

特に Y_1 と $\sqrt{Z} \cos \Theta$ の分布も等しい事に注意すれば, 指数分布, 一様分布がともに実現可能であるので例えば

```
#define lambda 2
double normal(){ return sqrt( lambda*exp() ) * cos( 2*M_PI*urand() ); }
```

によって正規分布のサンプルが得られる. 実際上は関数呼び出しが重い可能性があるので

```
double normal(){
    return sqrt( -2*log( (rand()+1.0)/(RANDMAX+1.0) ) )
           * cos( 2*M_PI*(double)rand()/RANDMAX );
}
```

の様にした方が良さそう¹.

¹ もう一つ指数関数の引数として 0 を渡さないように調整している.