

C Gのための確率・統計入門

mebiusbox software

2018 年 9 月 13 日

1 はじめに

コンピュータグラフィックスでは、様々なところで確率と統計が使用されています。例えば、レイトレーシングによるレンダリングでは、モンテカルロ法を使って光輸送のシミュレーションを行い、計算を高速に行えます。モンテカルロ法は乱数を使って数値積分を行う方法です。また、物理ベースレンダリングでよく使用されるマイクロファセット反射モデルでは、マイクロファセットの分布に確率が使用されています。他にもシャドウマップの影生成手法には確率を使ってソフトシャドウをシミュレーションするバリエーションシャドウマッピングがあります。このように確率と統計はコンピュータグラフィックにおいて重要な学問ではあるのですが、扱っている範囲が広く、個人的には難しい印象でしたし、今でもそう感じています。確率・統計は用語が多く、しかも似たような名前も少なくないので、混乱しがちです。そのため、用語は出来るだけ絞っています。目標はモンテカルロ法によって数値積分できるようになることです。

1.1 数式処理ソフトウェア

数式の計算や、シミュレーションのために Maxima と Scilab というソフトウェアを利用しています。本文中に各ソフトウェアのコマンドを記載しているところがあります。実行するにはインストールする必要があります。下記からダウンロードできます。

- Maxima (<http://maxima.sourceforge.net/>)
- Scilab (<https://www.scilab.org/>)

2 確率

まずは高校で学ぶ確率についてです。これは数学者ラプラスが定義した確率論に基づいています。確率の例として必ず出てくるサイコロを使いましょう。よくある 6 面で 1 から 6 の目があるものです。24 面とか 32 面とかではありませんよ。あとサイコロは 1 個です。

2.1 試行, 標本, 事象

今、サイコロを 1 回振りました。出た目は「1」でした。もう一度振ってみると「2」でした。細工をしない限り 1 回目と 2 回目の出る目は偶然によって決まります。このように同じ条件で何度でもサイコロを振ることを**試行**といいます。試行によって起こりえる結果はサイコロの目である「1, 2, 3, 4, 5, 6」のことで、1 つひとつを**標本**といいます。そして、すべての標本（ここでは 1, 2, 3, 4, 5, 6 の 6 つ）を**全事象**といいます。さらに全事象から取り出した標本の集まりを**事象**といいます。事象は 1 つの標本でもいいし、全部（すなわち全事象）でも構いません。

全事象は**標本空間**ともいいます。また、標本が 1 つだけの事象を**根元事象**といいます。

2.2 集合, 要素

互いに識別できる「もの」の集まりを、**集合**といいます。たとえば、1 から 5 までの自然数 1, 2, 3, 4, 5 の集まりは集合です。この場合、{1, 2, 3, 4, 5}、あるいは $A = \{1, 2, 3, 4, 5\}$ と表します。1 つひとつを**要素**といいます。要素をまったくもたない集合のことを**空集合**といい、 ϕ で表されます。ある集合 A の要素の個数は $n(A)$ と表します。

2.3 共通部分と和集合

集合 B の要素がすべて集合 A の要素であるとき、 B は A の**部分集合**であるといいます。そして $B \subset A$, または $A \supset B$ と表します。

一般に、集合 U を全体の集合とし、この中で部分集合を考えるとき、 U を**全体集合**といいます。部分集合 A と B の要素を合わせた集合を**和集合**といい、 $A \cup B$ と表します。また、部分集合 A と B のどちらにも入っている共通の要素を**共通要素**といい、 $A \cap B$ と表します。全体集合 U のうち、部分集合 A ではない要素を A の**補集合**といい、 \bar{A} と表します。

全体集合 U 、部分集合 A 、補集合 \bar{A} は次のような関係になります。

$$A \cup \bar{A} = U, \quad A \cap \bar{A} = \phi$$

2.4 確率の定義

2.4.1 数学的確率

標本空間 U が、 n 個の根元事象で構成されていて、どの2つの根元事象も「重複して起こらず」、また、どの根元事象の起こることも、「同様に確からしい」とします。ある事象 A が a 個の根元事象から成り立つとき、つまり、 A の起こる場合が a 通りであるとき、

$$p = \frac{a}{n},$$

のことを、事象 A の起こる確率といい、 $P(A)$ と表します。これを**数学的確率**と呼びます。

確率 $P(A) = 0$ のとき、事象 A は絶対に起こりません。 $P(A) = 1$ のときは、必ず起こります。サイコロの例で1の目が出る確率は、事象 $A = \{1\}$ で、全事象 $U = \{1, 2, 3, 4, 5, 6\}$ なので、確率は次のようになります。

$$P(A) = \frac{1}{6}.$$

2.4.2 統計的確率

ある試行の全事象を Ω と表します。この試行の1つの事象を A とします。全事象の個数を $n(\Omega)$ 、事象 A の個数を $n(A)$ とすると、事象 A が起こる確率 $P(A)$ は次のようになります。

$$P(A) = \frac{n(A)}{n(\Omega)}$$

この確率は**統計的確率**といいます。数学的確率は同様に確からしいという状況のもとに考えるのに対して、統計的確率では実際に試行した結果をもとに考えます。たとえば、サイコロに細工をして1の目が他の目よりも出やすくなっているものは同様に確からしいとは言えません。この場合、1の目が出る確率は $\frac{1}{6}$ ではないので、実際に試行した結果をもとに確率を算出します。

ある試行を n 回観測します。このとき事象 A が起きた回数を r としたとき、

$$\frac{r}{n},$$

これを**相対頻度**といいます。相対頻度が、 n を限りなく大きくしていったときに、一定の値 α に限りなく近づくとすれば、事象 A の確率 $P(A)$ を

$$P(A) = \alpha$$

と定められます。 **大数の法則** (Law of Large Numbers, またはベルヌーイの定理) によって、このことが証明されています。これはモンテカルロ法の根源となる定理です。

2.5 排反

事象 A と事象 B が同時に起こらないとき、 A と B は**排反である**といいます。たとえばサイコロの場合、事象 A が偶数 $\{2, 4, 6\}$ で、事象 B が奇数 $\{1, 3, 5\}$ だと排反であるといえます。しかし、事象 A が奇数 $\{1, 3, 5\}$ で、事象 B が3以下 $\{1, 2, 3\}$ の場合は**排反ではない**ということになります。

2.6 加法定理

事象 A と事象 B が排反であるならば、1回の試行で事象 A と事象 B が起こる確率は以下ようになります。

$$P(A \cup B) = P(A) + P(B)$$

2.7 条件付き確率

事象 B が事象 A が起きたことで確率が変わる場合、事象 B のことを**条件付き確率**といいます。この条件付き確率は $P_A(B)$ または $P(B|A)$ と表します。

例えば、10本のくじがあり、3本の当たりくじがあったとします。先に太郎が1本引き、次に次郎が1本引くとします。太郎が当たりくじを引く確率は10本中の3本なので、 $\frac{3}{10}$ です。次郎は、もし太郎が当たりくじを引いていた残り9本中の2本となるので、当たりくじを引く確率は $\frac{2}{9}$ です。太郎が当たりくじを引いていなかったら9本中の3本となるので、当たりくじを引く確率は $\frac{3}{9}$ となります。つまり、次郎の当たりくじを引く確率は条件付き確率ということになります。

2.8 乗法定理

条件付き確率のときの当たりくじにおいて、太郎と次郎が二人とも当たりくじを引く確率は次のようになります。

$$P(A \cap B) = P(A)P_A(B)$$

3 統計

統計には、一部のデータから全体を推測する**推測統計**と、全体のデータから役に立つ情報を得る**記述統計**があります。モンテカルロ法は確率によって真値を推測するため、推測統計に分類されます。

3.1 データ、階級、度数、度数分布

確率という試行を統計では観測（実験や調査）といいます。観測で得られた観測値を**データ**といいます。データをいくつかの**階級**に分けて考えたとき、各階級ごとのデータの数を**度数**といいます。この度数を縦軸に、階級を横軸にして表示したグラフを**ヒストグラム**といいます。ヒストグラムによってデータがどのように分布されているのかが分かりやすくなります。

3.2 確率変数

試行の結果に値が変わるものを**確率変数**といい、 X, Y などと大文字を用いて表されます。ある試行の結果 ω によって X の値が決まる場合、 $X(\omega)$ と書くことができます。確率変数には数値が入ります。たとえばコインの場合、標本空間は表と裏ですが、表が 1、裏が 0 のように具体的な数値が入ります。

3.3 確率分布

たとえば、サイコロを振ったときの出る目を確率変数とすると、確率変数に入る値は $\{1, 2, 3, 4, 5, 6\}$ のどれかになります。それぞれ出る頻度は確率で表されます。ある確率変数 X に対して、値とその確率をまとめたものを**確率分布**といい、それを表にしたものを**確率分布表**といいます。例えば、サイコロの場合、

X の値	1	2	3	4	5	6
確率	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

サイコロに細工をして 1 の目だけ出るようにした場合は次のようになります。

X の値	1	2	3	4	5	6
確率	1	0	0	0	0	0

確率分布の特徴は「期待値（平均）」と「分散」でわかります。

3.4 期待値（平均値）

特徴を表す代表値に**期待値**というのがあります。ある確率変数 X の期待値は $E[X]$ と書きます。

確率変数の取る値を x_1, x_2, \dots, x_n とし、 x_1 が出る確率を p_1 だとします。表に表すと次のようになります。

X の値	x_1	x_2	\dots	x_n
確率	p_1	p_2	\dots	p_n

期待値を数式で表すと

$$E[X] = \sum_{i=1}^n x_i \cdot p_i.$$

この期待値は平均とも考えられます。

確率変数 X の値に対応した確率の合計は 1 になります。

$$\sum_{i=1}^n p_i = 1.$$

3.5 分散、偏差、標準偏差

確率変数 X と期待値 $E[X]$ との差を**偏差**といいます。

$$|x_1 - E[X]|, |x_2 - E[X]|, \dots, |x_n - E[X]|.$$

絶対値を含むと取り扱いがしづらいので2乗します。2乗したものの総和を**分散**といいます。確率変数 X の分散は $V[X]$ と書きます。

$$V[X] = \sum_{i=1}^n (x_i \cdot p_i)^2.$$

期待値を使って表すと

$$V[X] = E[(X - E[X])^2].$$

次のように変形することができます。

$$V[X] = E[X^2 - E[X]^2].$$

分散は2乗しているため、平方根を取ることで確率変数 X の値との単位を一致させます。これを**標準偏差**といい、 $S[X]$ で表します。

$$S[X] = \sqrt{V[X]}$$

3.6 確率密度

これまでの確率は離散的な標本を対象にしていました。今度は連続的な標本を対象にします。たとえば、高校のあるクラスの身長を考えてみます。160cm や 160.5cm とか 175.4cm とか様々です。あるクラスの人数が 30 人ならば、最大で 30 標本あることになりますが、標本としては任意の値が取れることになります。この場合、例えば 160cm～165cm の身長の人がどれくらいの確率なのかというのがわかるとよさそうです。このように連続的な値に対する確率を表すときは**確率密度関数** (Probability Density Function: pdf) を使います。ある区間 $[a, b]$ の確率は、

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

ここで $f(x)$ が確率密度関数です。

確率密度関数はすべての領域で積分したときに 1 になっていなければなりません。

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

ある値 x までの確率密度を知りたい場合は $-\infty$ から x までの領域で定積分します。これを**累積分布関数** (Cumulative Distribution Function: cdf) といいます。

$$cdf(x) = \int_{-\infty}^x f(x) dx = P(X \leq x).$$

累積分布関数を微分すると、確率密度関数になります。

$$f(x) = \frac{d}{dx} cdf(x).$$

連続的な確率変数 X の期待値 ($E[X] = \mu$)、分散 ($V[X] = \sigma^2$) は次のように積分で求めます。

$$\mu = \int_a^b x f(x) dx, \quad (1)$$

$$\sigma^2 = \int_a^b (x - \mu)^2 f(x) dx. \quad (2)$$

標準偏差 ($S[X] = \sigma$) は次のようになります.

$$\sigma = \sqrt{\sigma^2}.$$

3.7 様々な確率分布

3.7.1 ベルヌーイ分布 (Bernoulli distribution)

コイン投げを考えます. 標本空間は {表, 裏} でそれぞれ {1, 0}, 表の確率を p , 裏の確率を $1-p$ とします. このように起こり得る結果が2通りで確率が等しくない試行を表す確率分布が**ベルヌーイ分布**です. このときの試行を**ベルヌーイ試行**ともいいます.

ベルヌーイ分布の期待値は p で, 分散は $p(1-p)$ です.

3.7.2 二項分布 (Binomial distribution)

ベルヌーイ試行を n 回行い, それが成功する確率を p としたときの確率分布を**二項分布**といいます. $Bi(n, p)$ と表すことがあります. ベルヌーイ分布は $Bi(1, p)$ のことです.

二項分布の期待値は np で, 分散は $np(1-p)$ です.

3.7.3 一様分布 (Uniform distribution)

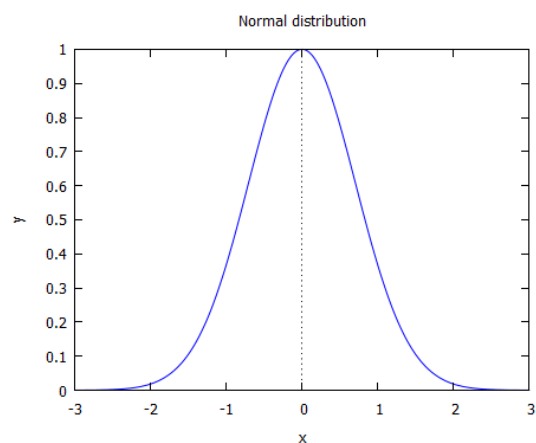
確率変数 X のとる値がすべて同じ確率のとき, この確率分布を**一様分布**といいます. 連続的な確率変数の場合, 確率密度関数が $f(x) = 1$ であることを意味しています. この場合, 期待値と分散は次のようになります.

$$\mu = \int_a^b x f(x) dx = \int_a^b x dx = \frac{1}{2}, \quad (3)$$

$$\sigma^2 = \int_a^b \left(x - \frac{1}{2}\right)^2 f(x) dx = \int_a^b \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12}. \quad (4)$$

3.7.4 正規分布 (Normal distribution)

ガウス分布 (Gaussian distribution) ともいわれています. 下図のような形をしています.



この曲線をベル曲線といいます。この図は `maxima` でプロットしたもので、次のようなコマンドです。

```
1 plot2d(%e^(-x^2), [x, -3, 3], [ylabel, "y"], [title, "Normal distribution"]);
```

ベル曲線を数式で表すと

$$e^{-x^2}.$$

正規分布の確率密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$\frac{1}{\sqrt{2\pi}\sigma}$ は以下を満たすための正規化定数です。

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

これは

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \sqrt{2\pi}\sigma,$$

から来ています。正規分布の期待値は

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu.$$

分散は

$$V[X] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2.$$

つまり、期待値と分散は指定した μ , σ^2 そのものとなります。正規分布の確率密度関数を、期待値（平均） μ 、分散 σ^2 から $N(\mu, \sigma^2)$ と表すことがあります。

3.7.5 その他の分布

指数分布, ポアソン分布, ガンマ分布, ベータ分布, χ^2 (カイ二乗) 分布, コーシー分布, 対数分布, パレート分布, ワイブル分布, t 分布, etc.

4 モンテカルロ法

モンテカルロ法は乱数を使って数値計算を行う方法です。コンピュータグラフィックではレイトレーシングでよく使われています。その特徴に、高次元積分でも適用できる、計算が早いといったメリットがありますが、真値への収束が遅いといった欠点があります。

4.1 円周率を求めてみる

よくある例として円周率 π を求めてみます。半径が 1 の円を考えます。この円の面積は $S = r^2\pi$ です。また、円を囲む正方形は一辺の長さが $2r$ となります。この正方形内に無作為に N 個の点を描いたとき、点が円の内部に入っている割合は、円の面積と正方形の面積の比になると推定されます。大数の法則によれば、 N を十分に大きくすると、実際の比と考えることができます。

$$\frac{r^2\pi}{(2r)^2} = \frac{\pi}{4}$$

試しにやってみます。以下は Scilab のコマンドです。

```
1 n = 10;  
2 x = rand(1,n)*2-1;  
3 y = rand(1,n)*2-1;  
4 r = x.*x + y.*y;  
5 ans=length(find(r<=1))/n*4  
6 disp(ans);
```

n を変えて計算した結果は次の通りです。

n	answer
10	4
1000	3.128
1000000	3.14118

n の回数を増やせば π に近づいていることが分かります。また、乱数のため実行する度に結果が変わります。

4.2 モンテカルロ法による数値積分の基礎

例えば 1 次元の積分で考えてみます。

$$I = \int_a^b f(x) dx.$$

この値を求めるために、区間 $[a, b]$ 上での一様乱数 v を発生させます。このような乱数は区間 $[0, 1]$ の一様乱数 u から $v = a + u(b - a)$ で求められます。次に

$$X = (b - a)f(v),$$

を計算してみます。確率変数 X の期待値 $E[X]$ は v の確率密度関数が

$$p(v) = \begin{cases} \frac{1}{b-a} & (v \in [a, b]) \\ 0 & (v \notin [a, b]) \end{cases},$$

なので、

$$E[X] = \int_a^b X p(v) dv = \int_a^b (b - a)f(v) \frac{1}{b-a} dv = \int_a^b f(x) dx = I,$$

となって、求める積分の値に等しくなります。

乱数 v を独立に N 個発生させて、それぞれに対して $X_i = (b - a)f(v_i)$ を計算して、それらの N 個の X_i の平均をとると次のようになります。

$$\hat{X}_N = \frac{1}{N} \sum_{i=1}^N X_i = \frac{b-a}{N} \sum_{i=1}^N f(v_i).$$

その期待値と分散は次のようになります。

$$\begin{aligned} E[\hat{X}_N] &= \frac{b-a}{N} \sum_{i=1}^N f(v_i) \\ &= (b-a) \cdot E[f(v_i)] \\ &= (b-a) \cdot \int_a^b f(v) p(v) dv \\ &= (b-a) \cdot \int_a^b f(v) \frac{1}{b-a} dv \\ &= \int_a^b f(v) dv = I. \end{aligned}$$

$$\begin{aligned} V[\hat{X}_N] &= V\left[\frac{1}{N} \sum_{i=1}^N X_i\right] \\ &= \frac{1}{N^2} V\left[\sum_{i=1}^N X_i\right] \\ &= \frac{1}{N^2} \sum_{i=1}^N V[X_i] \\ &= \frac{1}{N} \sigma^2. \end{aligned}$$

まとめると

$$E[\hat{X}_N] = I, \quad V[\hat{X}_N] = \frac{1}{N} \sigma^2.$$

標準偏差 $\sigma[\hat{X}_N]$ は次のようになります。

$$\sigma[\hat{X}_N] = \frac{1}{\sqrt{N}} \sigma[f(X)]$$

4.3 実際にモンテカルロ法で積分してみる

次の積分を求めてみます。

$$I = \int_0^2 x^2 dx$$

モンテカルロ法による一様乱数 u での推定値は次のようになります。

$$\hat{I} = \frac{2}{N} \sum_{i=1}^N f(X_i)$$

確率変数 X は次のようになります。

$$X = 2u$$

これでシミュレーションしてみます。以下は `scilab` のコードです。

```
1 N=[10, 1000, 1000000];
2 for i=1:length(N)
3     x=2*rand(1,N(i));
4     ans=2*mean(x.*x);
5     disp(ans);
6 end
```

結果は次のとおりです。乱数を使っているため、実行するたびに答えが変わるので注意してください。

n	answer
10	3.0309054
1000	2.7869074
1000000	2.6679645

今度は `maxima` を使って積分を計算してみます。

```
1 integrate(x^2,x,0,2);
```

$\frac{8}{3}$ と表示されました。 `float(8/3)` と入力すると `2.666666666666667` でしたのでほぼ真値になっています。

4.4 分散を減らす方法

モンテカルロ法の弱点は収束が遅いことです。そのため、サンプリング数を大きくしないと真値に近づきません。収束を早くする、つまり分散値を 0 に近づけることが重要になります。

4.4.1 層化サンプリング (Stratified sampling)

積分区間を小区間に分けて、小区間内で無作為に1つのサンプルを抽出します。

4.1 の円周率を、層化サンプリングで求めてみます。ここでは正方形を $n \times n$ 個に分割します。

以下は scilab のコマンドです。

```
1  n=10;
2  x=rand(n,n);
3  y=rand(n,n);
4  r=zeros(n,n);
5  for i=1:n
6      for j=1:n
7          X=2*((i+x(i,j))/n)-1;
8          Y=2*((j+y(i,j))/n)-1;
9          r(i,j) = X*X+Y*Y;
10     end
11 end
12 ans=4*length(find(z<=1))/(n*n);
13 disp(ans);
```

実行結果は次の通りです。

n	answer
10	2.76
100	3.1264
1000	3.141336

通常のモンテカルロ法よりも少ないサンプルで真値に近づいていることがわかります。実装が比較的簡単ではありますが、次元が増えるとサンプル数も指数的に増加していきますので、高次元では使用しにくい問題があります。

4.4.2 重点サンプリング (Importance sampling)

被積分関数 $f(x)$ の値の大きさに比例するような確率分布の乱数を使うことで分散を小さくすることができます。

一様分布ではない確率密度関数 $p(x)$ の確率変数を X とすると

$$I = \int_a^b \frac{f(x)}{p(x)} dx.$$

この推定値 \hat{X}_N は次のようになります。

$$\hat{X}_N = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)}.$$

次に期待値 \hat{X}_N は次のようになります。

$$\begin{aligned}
E[\hat{X}_N] &= E\left[\frac{1}{N}\sum_{i=1}^N \frac{f(X_i)}{p(X_i)}\right], \\
&= \frac{1}{N}\sum_{i=1}^N E\left[\frac{f(X_i)}{p(X_i)}\right], \\
&= \frac{1}{N}\sum_{i=1}^N \int_a^b \frac{f(x)}{p(x)} p(x) dx, \\
&= \frac{1}{N}\sum_{i=1}^N \int_a^b f(x) dx, \\
&= \int_a^b f(x) dx = I
\end{aligned}$$

分散 $V[\hat{X}_N]$ は次のようになります。

$$V[\hat{X}_N] = \frac{1}{\sqrt{N}} \sigma \left[\frac{f(X)}{p(X)} \right].$$

次の数式を重点サンプリングで求めてみます。

$$\int_0^2 x^2 dx$$

区間 $[0, 2]$ の確率密度関数 $p(x)$ を考えてみます。 $p(x)$ は x に比例するとします。数式で表すと

$$P(0 \leq x \leq 2) = \int_0^2 C p(x) dx.$$

C は定数で傾きになります。確率密度関数は積分すると 1 になることがわかっていますので

$$\int_0^2 p(x) dx = 1,$$

を満たさなければなりません。これを満たす定数 C を求めます。 $p = Cx$ とすると

$$\int_0^2 Cx dx = \left[\frac{1}{2} Cx^2 \right]_0^2 = 2C = 1,$$

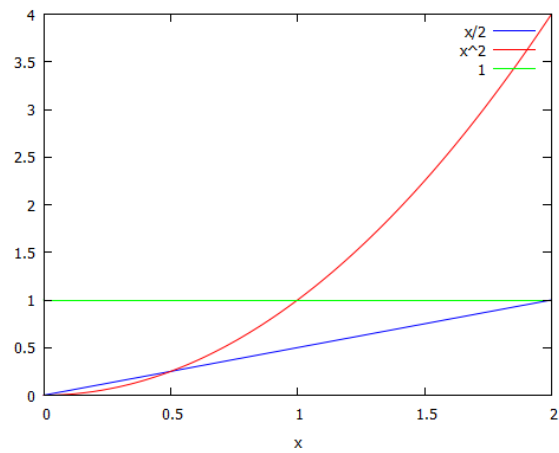
よって

$$C = \frac{1}{2},$$

だから

$$p(x) = \frac{x}{2}.$$

被積分関数は x^2 で、確率密度関数は $\frac{x}{2}$ 、区間 $[0, 2]$ を図にしてみると次のようになります。



確率密度が一様分布（緑）より，求めた確率密度関数（青）の方が被積分関数（赤）に形状が似ています．このように被積分関数の形状に近い確率密度関数を決めることが重点サンプリングの重要なことです．

確率密度関数 $p(x)$ の累積分布関数 $cdf(x)$ は

$$cdf(x) = \int_0^x \frac{x}{2} dx = \frac{1}{4}x^2 \quad (0 \leq x \leq 2)$$

となります．例えば $cdf(1.0) = \frac{1}{4}$ となり，これは x が 1.0 のとき，確率が 25% になるということを表しています．今，一様乱数 u を使って確率密度が $\frac{1}{4}x^2$ となる関数 $g(u)$ を求めたいわけです．これは $g(0.25) = 1.0$ となるはずです．

$$g(cdf(u)) = u$$

なので

$$g(u) = cdf^{-1}(u)$$

$cdf(x) = \frac{1}{4}x^2$ なので，

$$u = \frac{1}{4}x^2$$

$$4u = x^2$$

$$x = \sqrt{4u}$$

このようにある分布の確率密度関数を逆関数を使って求める方法を**逆関数法**といいます．

求めた確率密度関数を使ってシミュレーションしてみます．以下は `scilab` のコードです．

```

1 function ret = pdf(x)
2     ret = 0.5*x;
3 endfunction
4
5 N=[10, 1000, 1000000];
6 for i=1:length(N)
7     x=sqrt(4*rand(1,N(i)));
8     y=(x.*x)./pdf(x);
9     ans=mean(y);
10    disp(ans);
11 end

```

結果は次の通りです.

n	answer
10	3.0600588
1000	2.6592209
1000000	2.6657773

重点サンプリングにおいて, 理想的な確率密度関数は $p(x) = \frac{f(x)}{I}$ です. すでに I は $8/3$, $f(x) = x^2$ と分かっているので

$$p(x) = \frac{3}{8}x^2.$$

逆関数法を使って, 先ほどの確率密度に従う確率密度関数を計算します. まず, 確率密度関数から累積分布関数を求めると

$$cdf(x) = \int_0^x p(x) = \int_0^x \frac{3}{8}x^2 = \left[\frac{1}{8}x^3 \right]_0^x = \frac{1}{8}x^3.$$

次に逆関数を求めます.

$$cdf^{-1}(x) = 8x^{\frac{1}{3}}.$$

分散が 0 なので, サンプル数は 1 つで十分です. 実際にシミュレーションしてみます. 以下は scilab のコードです.

```

1 function ret = pdf(x)
2     ret = 3*x*x/8;
3 endfunction
4
5 N=[1];
6 for i=1:length(N)
7     x=(8*rand(1,N(i)))^(1/3);
8     y=(x.*x)./pdf(x);
9     ans=mean(y);
10    disp(ans);
11 end

```

結果は次のとおりです.

n	answer
1	2.6666667

5 乱数の生成

重点サンプリングするために、ある確率分布に従った乱数を作成する必要があります。既に、逆関数法を使ったやり方を説明しました。しかし、必ずしも逆関数法が使えるわけではありません。条件に累積分布関数が単調増加（または単調減少）である必要があります。

5.1 正規分布乱数

正規分布に従う乱数を作成する場合、その累積分布関数の逆関数を求めるのは簡単ではないので、中心極限定理を使った方法とボックス・ミュラー法を使って求めるやり方があります。まずは中心極限定理を使った方法を見ていきますが、その前に確率のところで少し出てきた大数の法則について、あらためて見ていきます。

5.1.1 大数の法則 (Law of large numbers)

コイン投げを考えます。標本空間は {表, 裏} で {1, 0} とし、10 回の試行をしました（2つのうちのいずれかの事象が発生する試行をベルヌーイ試行といいました）。ここで、確率変数 X_i は、 i 回目のコイン投げの結果です。表の出た回数（頻度）は、確率変数の和となります。

$$r = X_1 + X_2 + \dots + X_{10}$$

表の出た回数の割合 $\hat{p} = \frac{r}{10}$ は観測された成功率で、 $\hat{p} = 0, 0.1, 0.2, \dots$ となります。一般に、試行回数を n とするとき、 $\frac{r}{n}$ を相対頻度といいます。 r は確率変数で $n = 10$, $p = 0.5$ の二項分布 $Bi(10, 0.5)$ に従います。期待値と分散は二項分布の計算式を使うとそれぞれ

$$E[r] = np = 5 \quad (5)$$

$$V[r] = np(1-p) = 2.5, \quad (6)$$

となります。相対頻度の期待値と分散は次のとおりです。

$$E\left[\frac{r}{n}\right] = p = 0.5 \quad (7)$$

$$V\left[\frac{r}{n}\right] = \frac{p(1-p)}{n} = 0.025. \quad (8)$$

ここで、 $p = 0.5$ は真の成功率です。観測された成功率とその確率を表にまとめてみます。二項分布の確率は $f(x) = {}_nC_x p^x (1-p)^{n-x}$ です（ ${}_nC_r$ は組み合わせで $\frac{n!}{r!}$, ${}_nP_r$ は順列で $\frac{n!}{(n-r)!}$ です）。今回の場合は $f_i(x) = {}_{10}C_x (\frac{1}{2})^{10}$ となります。

観測された成功率	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
確率	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

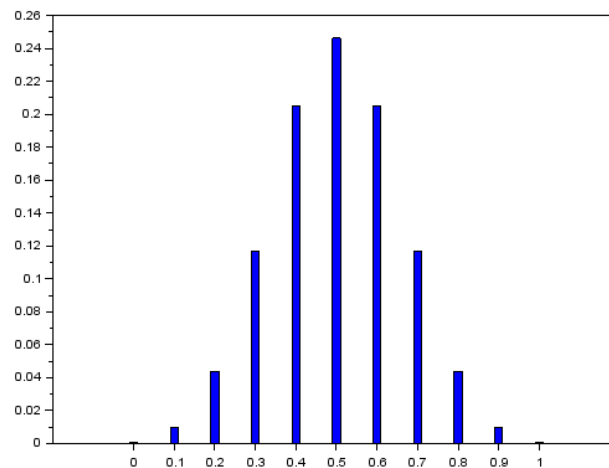
下記は scilab で確認したコードです。


```

1 function y = permutations(n,r)
2     y=factorial(n)/factorial(n-r);
3 endfunction
4
5 function y = combinations(n,r)
6     y = permutations(n,r)/factorial(r);
7 endfunction
8
9 n=10;
10 x=zeros(10);
11 for i=0:n
12     x(i+1) = combinations(10,i) * (1/2)^(10);
13 end
14 disp(x);

```

この表を見てもわかりづらいので、ヒストグラムにしたものが下図です。

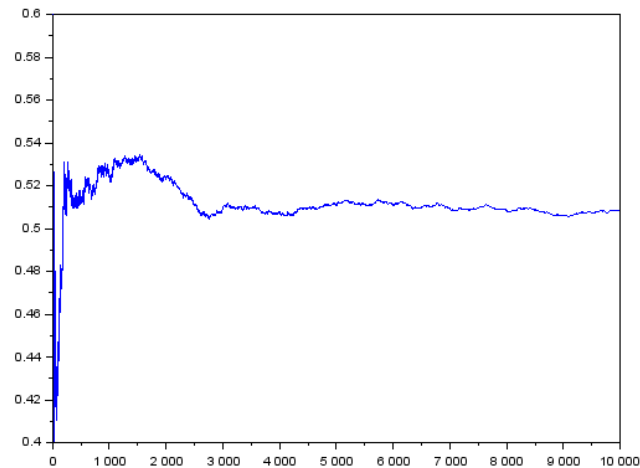


なんだかそれっぽい図ですね。

(ヒストグラムは先程の scilab のコードの最後に `bar(0:0.1:1,x,0.2);` を追加します)

ここで、コイン投げの回数 n を増やしていくと $\frac{i}{n}$ の期待値 $E[\frac{i}{n}] = p$ ，およびその周辺が発生する確率が上がっていき，最終的に真の成功率 $p = 0.5$ に集中し，相対頻度は真の成功率に収束していきます。これが**大数の法則**です。

$Bi(1,0.5)$ のベルヌーイ試行を繰り返していき，相対頻度の変化をグラフにしたものが下図です。真の成功率 $p = 0.5$ 近くに収束していくのがわかります。



以下は scilab のコードです.

```

1  n=10000;
2  x=bool2s(rand(1,n)<0.5);
3  y=cumsum(x)./(1:n);
4  plot(y);
5
6  a=gca();
7  a.data_bounds(:,2)=[0.4,0.6];

```

5.1.2 中心極限定理 (Central limit theorem)

期待値 μ , 分散 σ^2 である確率変数 X を考えます. この確率分布は一様分布かもしれないし, 二項分布かもしれないし, なんでも構いません. この試行を何度も繰り返していくと, 標本平均 (確率変数の平均) の確率分布はおおむね正規分布になるというのが**中心極限定理**です.

標本平均 S_n とすると

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

n が大きくなるにつれて S_n の分散は近似的に $N(\mu, \frac{\sigma^2}{n})$ となることがわかっています. また,

$$Z_n = \frac{S_n - \mu}{\sigma / \sqrt{n}},$$

と定義し, S_n を標準化すると n が大きくなるにつれて, 確率変数 Z_n は標準正規分布 $N(0, 1)$ に近づいていきます. 整理すると

$$Z = \frac{S - \mu}{\sigma},$$

ここで Z は**標準化変数**といいます.

5.1.3 二項分布による正規分布の近似

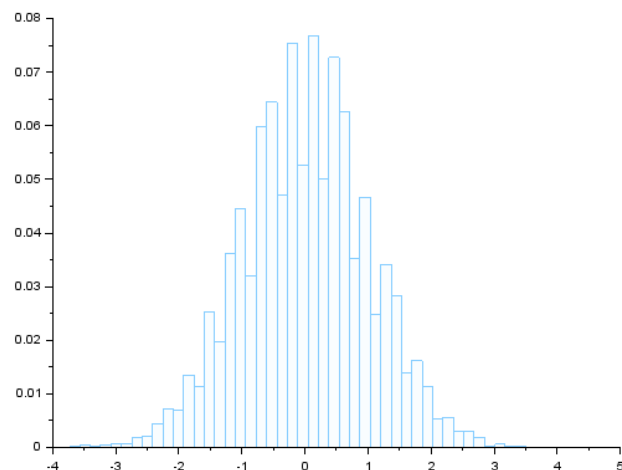
中心極限定理によれば、 n が大きいときに $Bi(n, p)$ は正規分布に近づくので、それを利用して正規分布乱数を作成することができます。

二項分布の期待値は $\mu = np$ 、分散は $\sigma^2 = np(1-p)$ なので、標準化変数に代入すると

$$Z = \frac{S - np}{\sqrt{np(1-p)}}.$$

$Bi(1000, 0.5)$ を scilab でシミュレーションしたコードと結果は次のとおりです。

```
1 p=0.5;
2 n=1000;
3 N=10000;
4 x=rand(N,n);
5 y=bool2s(x<p);
6 S=sum(y,"c");
7 m=n*p;
8 sigma=sqrt(n*p*(1-p));
9 Z=(S-m)/sigma;
10 histplot(50,Z,style=12);
```



5.1.4 正規分布乱数の生成

一様乱数から正規分布乱数を生成します。これも中心極限定理に基づきます。一様乱数の期待値と分散は $\mu = \frac{1}{2}$ 、 $\sigma^2 = \frac{1}{12}$ です。標準化変数は

$$z = \left(\sum_{i=1}^n X_i - \frac{n}{2} \right) \frac{1}{\sqrt{n/12}}.$$

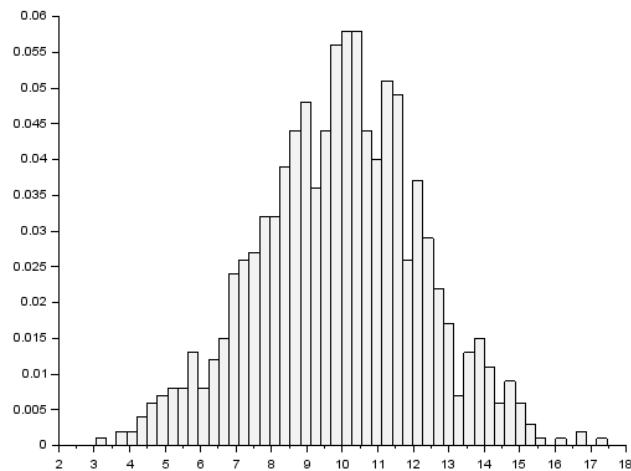
期待する正規分布 x の期待値 μ 、分散を σ^2 とすると、 x の標準化の逆変換 $x = \sigma z + \mu$ に代入して

$$x = \sigma \sqrt{\frac{12}{n}} \left(\sum_{i=1}^n X_i - \frac{n}{2} \right) + \mu.$$

これが、正規分布 $N(\mu, \sigma^2)$ に従う正規分布乱数です。

$N(10, 5)$ の正規分布を scilab でシミュレーションしたコードとヒストグラムは次のようになっています。

```
1 sigma=5;
2 mu=10;
3 n=1000;
4 N=1000;
5 x=rand(N,n);
6 S=sum(x, "c");
7 Z=sqrt(sigma)*sqrt(12/n)*(S-n/2)+mu;
8 histplot(50,Z);
```

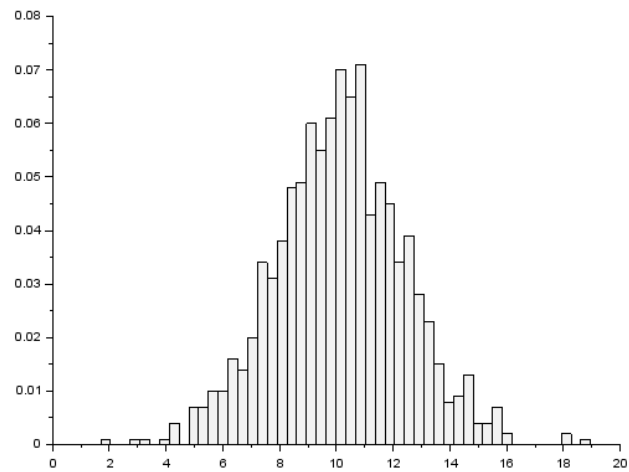


標準化変数で $n=12$ とすると式を簡単にすることができます。

$$x = \sigma \left(\sum_{i=1}^{12} X_i - 6 \right) + \mu.$$

これもシミュレーションした結果は次のとおりです。

```
1 sigma=5;
2 mu=10;
3 N=1000;
4 x=rand(N,12);
5 S=sum(x, "c");
6 Z=sqrt(sigma)*(S-6)+mu;
7 histplot(50,Z);
```



5.1.5 ボックス=ミュラー法 (Box – Muller's method)

詳しい証明はここでは行いませんが，以下の式を使うと正規分布乱数を生成できます． Z_1, Z_2 どちらかを使います．

$$Z_1 = \sqrt{-2 \log X} \cos(2\pi Y) \quad (9)$$

$$Z_2 = \sqrt{-2 \log X} \sin(2\pi Y). \quad (10)$$

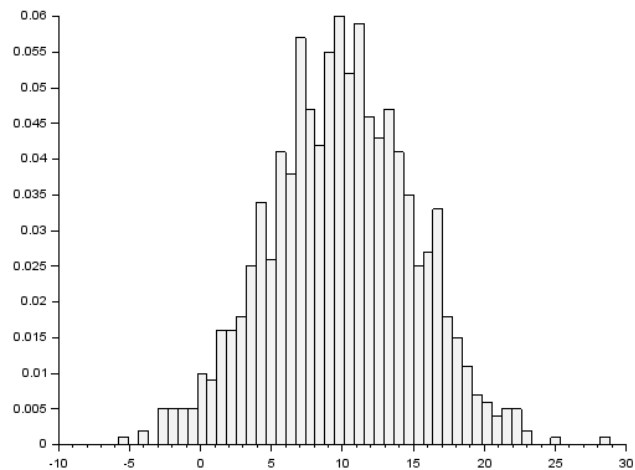
分散と平均を指定する場合は

$$Z_1 = \sigma \sqrt{-2 \log X} \cos(2\pi Y) + \mu \quad (11)$$

$$Z_2 = \sigma \sqrt{-2 \log X} \sin(2\pi Y) + \mu. \quad (12)$$

$N(10, 5)$ の正規分布を `scilab` でシミュレーションしたコードとヒストグラムは次のようになっています．

```
1  sigma=5;
2  mu=10
3  N=1000;
4  x=rand(1,N);
5  y=rand(1,N);
6  z=sigma*sqrt(-2*log(x)).*cos(2*%pi*y)+mu;
7  histplot(50,z);
```



6 最後に

モンテカルロ法による数値計算に必要と思われる確率と統計に関してまとめてみました。参考にした書籍は入門，超入門とかなので，まだまだ入り口部分の内容かもしれません。私自身は専門ではないので，間違いなど見つけましたらご連絡いただけると幸いです。確率密度関数では図ももう少し入れたらわかりやすいので，いずれ入れておこうと思います。

7 参考文献

- 郡山彬, 和泉澤正隆, 「確率と統計超入門」, 2001
- 東京大学教養学部統計学教室, 「統計学入門」, 2017
- リチャード・エルウィス著, 宮本寿代訳, 「マスペディア 1000」, 2016
- 涌井良幸, 「『数学』の公式・定理・決まりごとがまとめてわかる事典」, 2016
- 伊里正夫, 藤野和建, 「数値計算の常識」, 2010
- Shocker, 「モンテカルロ法 (Monte Carlo method)」 <http://rayspace.xyz/CG/contents/montecarlo/>
- Shocker, 「モンテカルロ法の分散低減手法」, http://rayspace.xyz/CG/contents/variance_reduction/
- Ushio, 「確率密度関数からモンテカルロ積分まで」, <https://qiita.com/Ushio/items/0040b3c74a480c46c80c>