８．確率論：基礎

- Random numbers
- Conditional probability 条件付確率
- Joint probability 結合確率
- Bayes’ theoremベイズの定理
- Marginal probability 周辺確率
- Posterior probability and prior probability 事後確率と事前確率
- Logical indexing of matrices

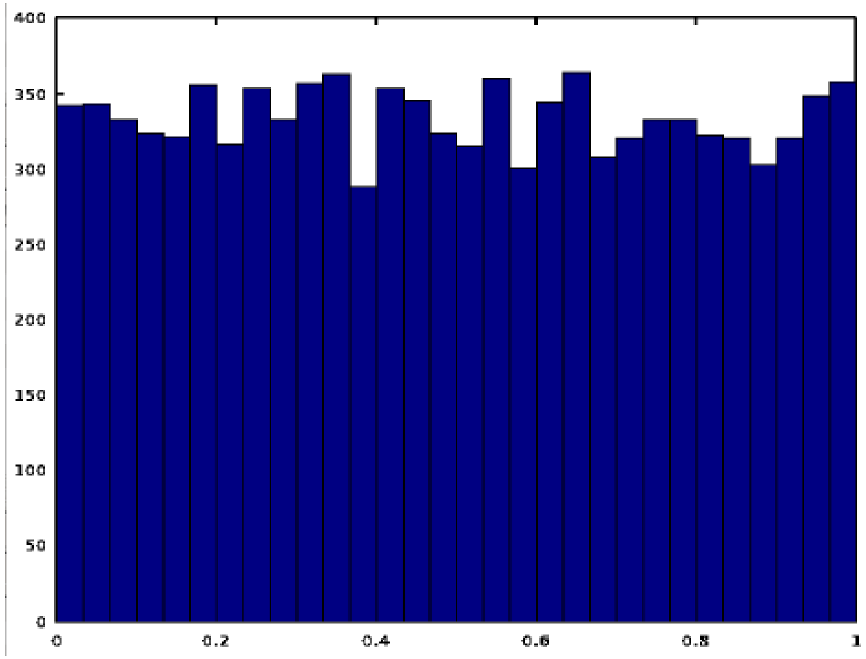
Random numbers

- コマンドrand(m,n)は0から1の間で均一な分布をもつランダムな数値をm×n行列に代入する.

Randを用いて10000個のランダムな数値を発生させ，それをhistコマンドを用いてヒストグラムにすると

```
>> hist(rand(10000,1),30)
```

０～１の数値が同じ頻度で発生する

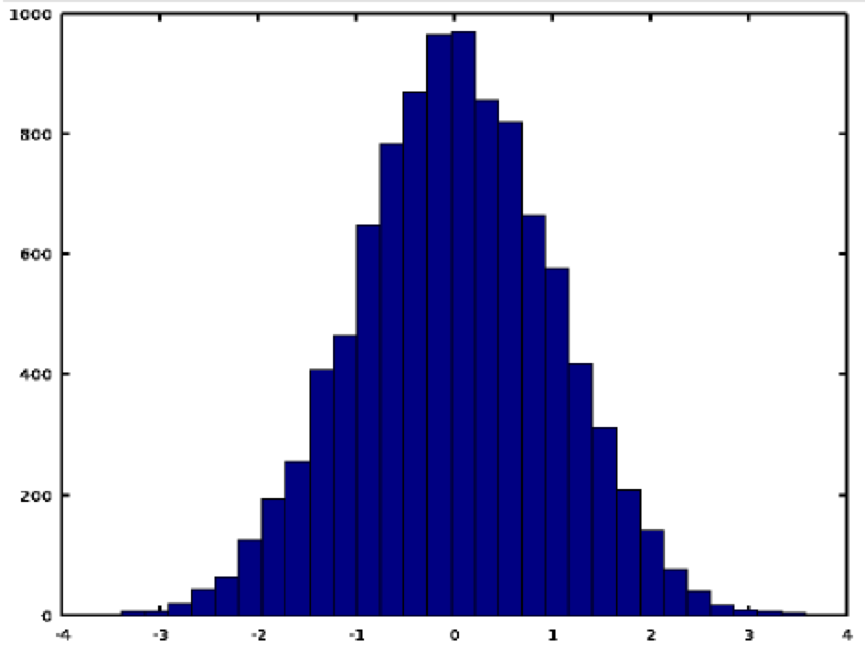


- コマンドrandn(m,n)は平均0，分散1の正規分布を持つランダムな数値をm×n行列に代入する.

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

μ：平均
σ：分散

```
>> hist(randn(10000,1),30)
```



Probabilistic theory 確率論

- 赤い箱と青い箱にそれぞれフルーツが入っている

2 apples, 6 oranges　　3 apples, 1 orange

1. まずはじめに，どちらかの箱を選ぶ.
2. その後選んだ箱からフルーツをランダムに１つ選ぶ.

- 赤い箱を選ぶ確率は40%, 青い箱を選ぶ確率は60%とする

例えば次のような問題はどのようにして解くのか？

- リンゴが取り出される確率
- リンゴが取り出された時，それが青い箱から取り出された確率

From 1st chapter in C. M. Bishop, Pattern Recongnition and Machine Learning

Conditional probability　条件付き確率

- 確率変数
  - B= どちらの箱が選ばれたかを指す；　B=r:赤い箱　B=b:青い箱
  - F= どちらのフルーツが選ばれたかを指す；　B=a: apple　B=o:orange
- それぞれの箱が選ばれる確率を表すと

$p(B=r)=4/10$　赤い箱を選ぶ確率40%　$p(B=r)=6/10$ 青い箱を選ぶ確率60%

- 条件付き確率 $p(A|B)$ すでにBが起きている状態でAが起きる確率

例）すでに赤い箱を選んでいる状態として、appleを選ぶ確率は？

赤い箱にはフルーツが8個あり、そのうちの２つがappleなので　$p(F=a|B=r)=2/8=1/4$

- 同様に残りの条件付き確率は

$$p(F=o|B=r)=3/4$$
$$p(F=a|B=b)=3/4$$
$$p(F=o|B=b)=1/4$$

Joint probability 結合確率

- 結合確率 p(A,B) BもAも両方が起きる確率

例）赤い箱が選ばれて、さらにorangeを選ぶ確率p(F=o,B=r)は？

ベイズの定理　$p(F=o,B=r)=p(F=o|B=r)p(B=r)$

$p(F=o|B=r)=3/4,\ p(B=r)=4/10$ ⇒ $p(F=o,B=r)=3/10$

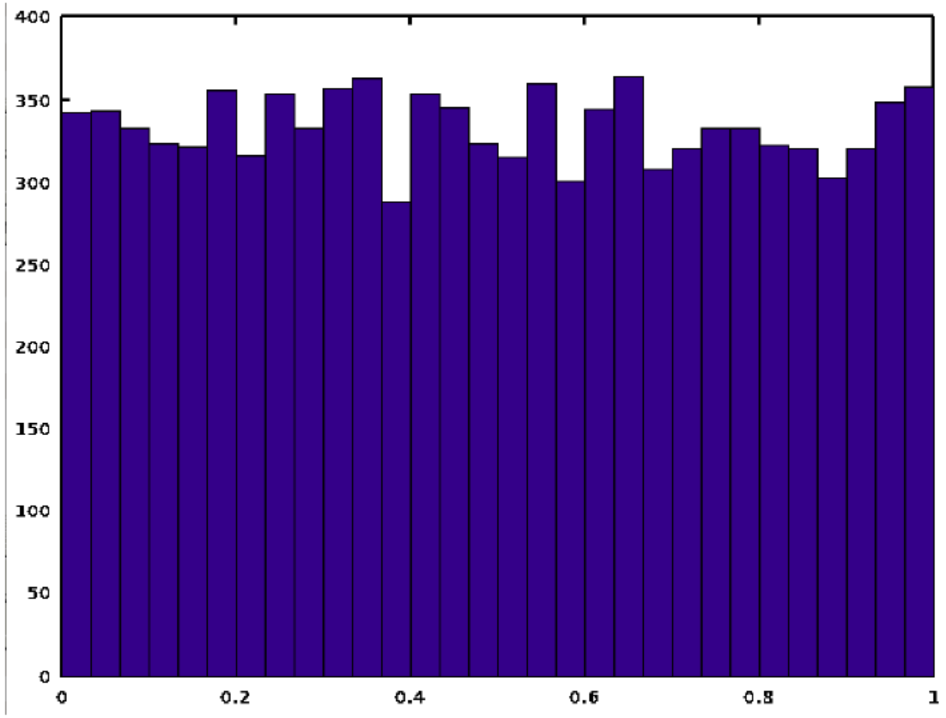
	$B=r$	$B=b$
$F=a$	1/10	9/20
$F=o$	3/10	3/20

ちなみに$p(F=o,B=r)=p(F=o)p(B=r)$が成り立つ場合には
2つの確率変数は互いに独立であるという

モンテカルロ推定

ランダムナンバーを使った確率計算

```
>> hist(rand(10000,1),30)
```



rand関数を使う
0から１までの数をランダムに発生させる.
その数値は0から１まで均一に表れる.

rand < 0.4になる回数は全体の4割
randで発生させた数が0.4未満になる確率は４０％

これを利用して，これまで議論していた箱とフルーツの問題を数値的に取り扱う

モンテカルロ法で確率を計算

箱を選んでフルーツを取り出すという試行を1万回繰り返して、どんな事象が何回起こったかを数える.

```
% box_fruit.m
num_bf = zeros(2,2);
for i=1:10000
  if rand(1,1) < 0.4, % red box (40%)
    if rand(1,1) < 2.0/8, % apple
      num_bf(1,1) += 1;
    else % orange
      num_bf(2,1) += 1;
    end
  else % blue box
    if rand(1,1) < 3.0/4, % apple
      num_bf(1,2) += 1;
    else % orange
      num_bf(2,2) += 1;
    end
  end
end
```

全ての成分が0の2×２行列num_bfを作成

For以下の内容を1万回行う

行列num_bf= 各事象が起きた回数をカウント

red box & apple	blue box & apple
red box & orange	blue box & orange

40%の確率で赤枠の内容を実行
(赤い箱が選ばれた場合)

2/8の確率でappleを選ぶ
行列num_bfの(1,1)成分の数を1増やす

6/8の確率でorangeを選ぶ
行列num_bfの(2,1)成分の数を1増やす

60%の確率で青枠の内容を実行
(青い箱が選ばれた場合)

3/4の確率でappleを選ぶ
行列num_bfの(1,2)成分の数を1増やす

1/4の確率でorangeを選ぶ
行列num_bfの(2,2)成分の数を1増やす

```
% box_fruit.m
num_bf = zeros(2,2);
for i=1:10000
  if rand(1,1) < 0.4, % red box (40%)
    if rand(1,1) < 2.0/8, % apple
      num_bf(1,1) += 1;
    else % orange
      num_bf(2,1) += 1;
    end
  else % blue box
    if rand(1,1) < 3.0/4, % apple
      num_bf(1,2) += 1;
    else % orange
      num_bf(2,2) += 1;
    end
  end
end
```

行列の各要素を全試行回数で割ってやると確率になる

```
>> box_fruit
>> num_bf/sum(sum(num_bf))
ans =
   0.10090   0.44660
   0.29740   0.15510
```

⇕

	B=r	B=b
F=a	1/10	9/20
F=o	3/10	3/20

Logical indexing of matrices　もっと効率的にモンテカルロ推定を行うには

```
>> B = rand(1,10000) < 0.4; % 1 for red box; 0 for blue box
>>
>> B(1:10)
ans =
    1    0    1    1    1    0    1    0    0    0
```

Bの各要素が0.4未満であれば１、そうでなければ０
１は赤い箱が選ばれ，0は青い箱が選ばれたことを示す

最初の10回どちらの箱が選ばれたかを抜き出すと・・・

```
>> Frnd = rand(1,10000);
>>
>> F(B==1) = Frnd(B==1) < 2/8;
>>
>> F(B==0) = Frnd(B==0) < 3/4;
>>
>> F(1:10)
ans =
    0    0    0    0    1    0    0    1    0    1
```

Bの内，値が１である要素の番号

B＝1の場合=>(赤い箱が選ばれた場合)
Frandの値が2/8より小の場合は値を１にする=>apple
2/8以上の場合は0とする　　=>orange

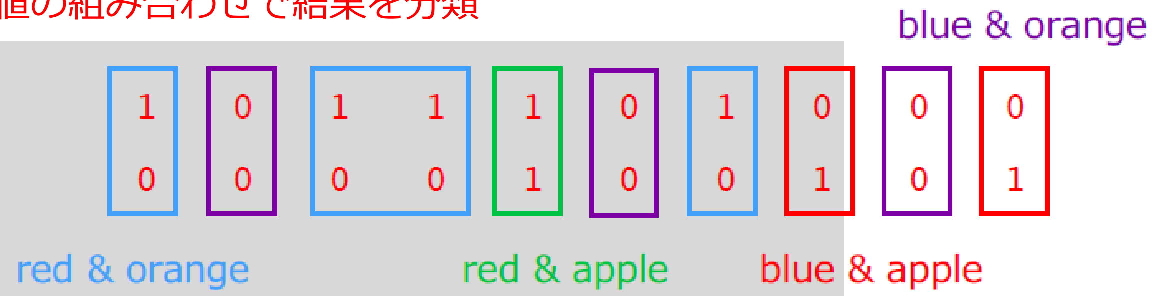
Bの内，値が0である要素の番号

B＝0の場合=>(青い箱が選ばれた場合)
Frandの値が3/4より小の場合は値を１にする=>apple
3/4以上の場合は0とする　　=>orange

最初の10回どちらのフルーツが選ばれたかの試行だけ抜き出すと・・・

```
>> sum(F==1&B==1)/10000
ans = 0.09570
>> sum(F==1&B==0)/10000
ans = 0.45430
>> sum(F==0&B==1)/10000
ans = 0.29890
>> sum(F==0&B==0)/10000
ans = 0.15110
```

ベクトルBは選ばれた箱，ベクトルFには選ばれたフルーツの情報が入っている
BとFの値の組み合わせで結果を分類



BとFの値のペアがそれぞれ何個あるかをカウントし，全試行回数で割れば確率になる

Logical indexing of matrices

```
>> A=rand(1,5)
A =

   0.0054027   0.0291497   0.4692166   0.0245417   0.6445969

>> B=A<0.4
B =

  1  1  0  1  0
```

Aの各要素が0.4未満であれば１、そうでなければ０

F(B==1)=Frnd(B==1)<2/8;

Bの要素の内，値が１であるものと同じ要素番号

例）

```
>>B(1:10)
ans=
       1    0    1    1    1    0    1    0    0    0
```

値が1の要素番号は 1, 3, 4, 5, 7

- marginal probability 周辺確率 p(A)

他の事象に関わりなくAが起きる確率

例） appleが選ばれる確率p(F=a) （Boxに関係なく）

p(F=a)=p(F=a,B=r)+p(F=a,B=b)　　1/10+9/20=11/20

- posterior probabilities 事後確率

ある情報を持った状態で変数の確率を求める

例） 情報：選ばれたフルーツはorangeである

そのorangeが取り出されたboxが赤いboxである確率は？

$$p(B=r|F=o)=\frac{p(B=r,F=o)}{p(F=o)}$$

$p(B=r,F=o)$=3/10　$p(F=o)$=9/20

$$p(B=r|F=o)=2/3$$

Exercise 8.1

- Logical indexing of matrices を使って確率p(F=a)のモンテカルロ推定を行え.
- Posterior probability 事後確率 p(B=r|F=o)をモンテカルロ推定で計算せよ