8. 確率論: 基礎

- Random numbers
- Conditional probability
- Joint probability
- Bayes’ theorem
- Marginal probability
- Posterior probability and prior probability
- Logical indexing of matrices

様々な事象に関して数値的（パソコン上で実験的）に確率を求める。

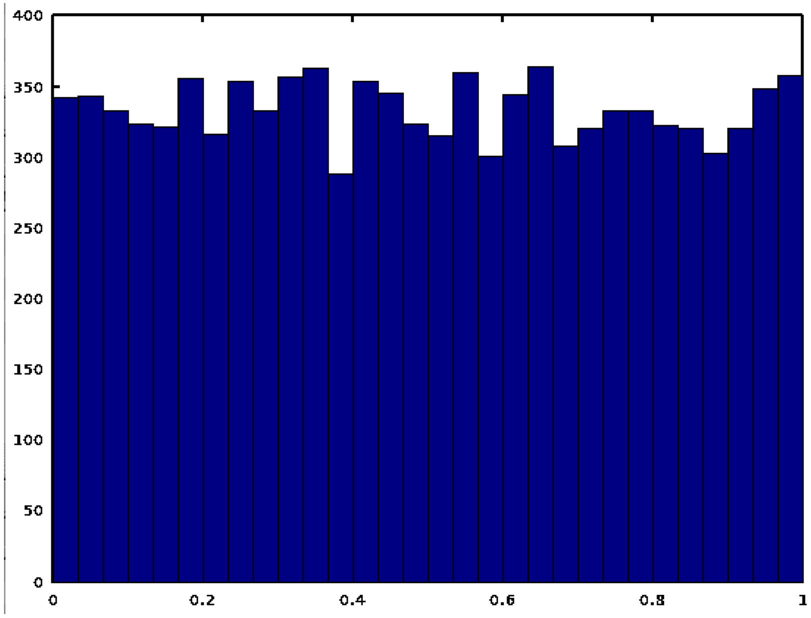
乱数 Random numbers

- `rand(m,n)`

要素が乱数の *m*×*n*行列を生成する。
各要素は0~1の間で均等な乱数である。

※ `hist` でヒストグラムを
作成できる。

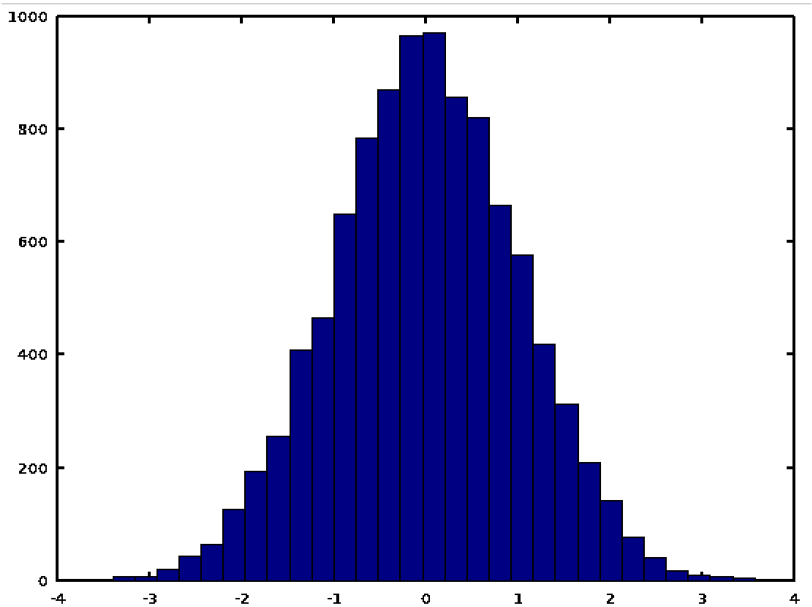
```
>> hist(rand(10000,1),30)
```



- `randn(m,n)`

要素が乱数の *m*×*n*行列を生成する。
ただし、各要素は分散が1、平均が0の
正規分布に従った乱数である。

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



確率論(1/4)

- 次の様に青リンゴとオレンジの入った赤い箱と青い箱を考える。

- まず箱をどちらか選び、次にフルーツを取る。
- 赤い箱は40%で選ばれ、青い箱は60%で選ばれると仮定する。
- そのとき、以下の問いにはどのように解答することができるか。
  - リンゴを取る確率は？
  - もし、リンゴを取ったことが分かっているとき、そのリンゴが青い箱から取られた確率は？

確率論(2/4)

- 確率変数
  - *B* (どちらの箱を選んだか); *B=r* (赤の箱を選んだ) , *B=b* (青の箱を選んだ)
  - *F* (どちらのフルーツを選んだか); *F=a* (リンゴ), *F=o* (オレンジ)
- それぞれの箱を選んだ場合の確率:

$$p(B=r)=4/10 \qquad p(B=b)=6/10$$

- 条件付き確率: 赤の箱を選んだ場合(確定)にリンゴを選ぶ確率は？
  - 箱の中のリンゴ(2個)の個数とフルーツ(8個)の総数の比で表される。

$$\underline{p(F=a|B=r)}=1/4$$

B=rが起きた後(条件下)でのF=aが起きる確率

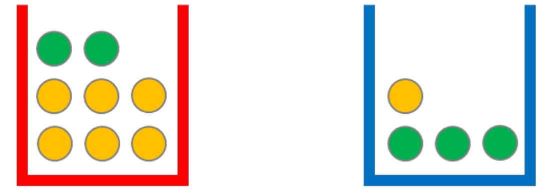
  - 同様に次の確率が求められる。

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

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

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

確率論(3/4)



2 apples, 6 oranges　3 apples, 1 orange

- 結合確率(同時確率)
  - 赤の箱を選び、かつ、オレンジを選ぶ確率は？
  - ベイズの定理によると次の式で与えられる。

$$p(F=a,\ B=r)=p(F=a|B=r)p(B=r)$$

かつ

	$B=r$	$B=b$
$F=a$	1/10(=4/10×2/8)	9/20
$F=o$	3/10	3/20

- 次の式が成り立つとき、一般に確率変数は互いに独立であると言われる。

⇒互いに独立な事象に対してのみベイズの定理*が成り立つ。

$$p(F=o, B=r)=p(F=o)p(B=r)$$

*ベイズの定理は機械学習の分野で良く用いられている。
例えば、迷惑メールフィルタなど。

モンテカルロ法

- モンテカルロ法を用いた確率の推定；乱数を用いて現象をシミュレーションすることで確率を数値的に推定する。
  - The following script first picks a box and then a piece of fruits randomly as is described above for, say, 10,000 trials; and counts the numbers of the cases (*B*, *F*) = (*r*, *a*), (*r*, *o*), (*b*, *a*), and (*b*, *o*), respectively

#と同様でコメントアウトを示す

```
% 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
```

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

各列内の和

全要素の和

⇕

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

行列の論理インデックス

- Octave上では行列で演算する方が効率が良い。

```
>> 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
```

“比較演算子”を用いて右辺に示す条件式を満たす要素を1、満たさない要素を0を代入

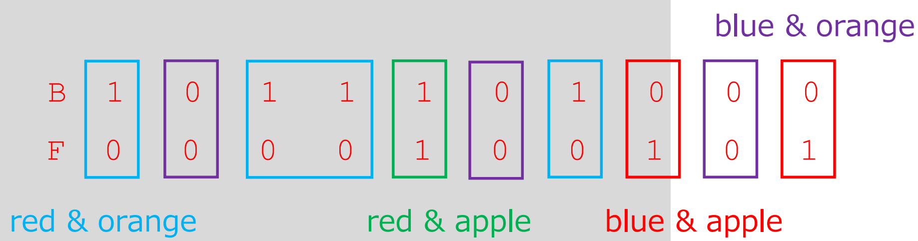
10000回の試行(要素)の内、最初の10回の結果を示す

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

行列Bにおける1(赤い箱)を示す要素(試行回数目)と一致する行列Fの要素を抽出(上のansの例なら1,3,4,5…番目)

行列Bにおける0(青い箱)を示す要素(試行回数目)と一致する行列Fの要素を抽出(上のansの例なら2,6,8,9…番目)

フルーツ選びにおいて、10000回の試行(要素)のうち、最初の10回の結果を表示



確率論(4/4)

- 一度の試行でリンゴを選ぶ確率は？
  - この種の確率は周辺確率と呼ばれる。
  - 答えは 11/20 で次の式で求められる。

$$p(F=a)=p(F=a,B=r)+p(F=a,B=b)$$

- もし、オレンジを選んだと言われたとき、そのオレンジが赤い箱から選ばれた確率は？
  - 次の式から 2/3 と求められる。

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

  - ある事象(今回はF)についての観測後に求める確率なので、事後確率と呼ばれる。
  - 一方で$p(B=r)$ のような確率は事前確率と呼ばれる。

Exercise 8.1

- １ページ前に示した確率$p(F=a)$ を行列を用いたモンテカルロ法で求めよ。
- 確率$p(B=r|F=o)$をモンテカルロ法で求めよ。

提出先：
東北大学ｲﾝﾀｰﾈｯﾄｽｸｰﾙ(ISTU)上で提出
もしくは
Email: hisashi.kino.a1＠tohoku.ac.jp
shimada@m.tohoku.ac.jp

〆切：
2018年7月13日(金)の午前8:50