情報数学II
第8回 確率・統計1

小野順貴
(ONO, Nobutaka)
東京都立大学 システムデザイン学部
情報科学科 教授
onono@tmu.ac.jp

小野分担講義予定

- 12/6(Wed) 開講
- 12/13(Wed) 開講
- 12/20(Wed) 開講
- 12/27(Wed) 補講期間のため休講
- 1/3 (Wed) 冬休みのため休講
- 1/10 (Wed) 開講
- 1/17 (Wed) 開講
- 1/24 (Wed) 開講
- 1/31 (Wed) 開講・期末試験（予定）

内容としては、確率・統計、最適化の基礎、フーリエ変換の基礎を行う予定

確率

- 数学的には公理的確率論などを通して厳密な定義が与えられるがここでは直感的に確率の概念を復習する。
- 事象の「起こりやすさ」を0から1の数値で定量的に表す指標。
  - 同じ条件で、多数、実験、ないしは観測を行う（試行という）とき、ある事象が発生する相対頻度の極限値
    （事象が起こった回数）／（試行の回数）
- 以下では事象Aが起こる確率を$P(A)$ で表す。

確率的な扱い

- 確率を考えるとき、事象が確定的か否かでは必ずしも関係ない。事象の観測に対する不確かさの度合いや事象の発生頻度の偏りの度合いを定量的に扱うために用いられる。
  - サイコロを振って何の目が出るのか、これは古典力学的には、サイコロを振る（落とす）ときの位置、初速度、重力、（下が机だとして）机の面の反発係数などで完全に決定され、確率的な要素はない。しかし、通常、これらの詳細な値は事前には知りえない。
  - 文章中に現れる文字には偏りがある（例えば、英語であれば、”e”の出現頻度が通常最も高い。）すでにかかれた文章は完全に確定的であり、不確定性はない。しかし、ある文章から無作為に選んだ文字が何であるかは、確率的に扱うことができる。

条件付き確率

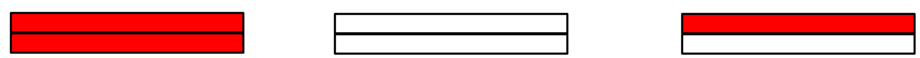
- 事象Bが起こるという条件のもとで、別の事象Aが起こる確率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

事象A, Bが共に起こる確率

事象Bが起こる確率

- 例題：両面赤(R)、両面白(W)、片面赤片面白(M)の3枚のカードがある。1枚を無作為に選び机に置くと、表面は赤であった。裏面が白である確率を求めよ。



- 解答：各カードの両面を区別し、便宜上1、2とすると、カードの表面は、R1, R2, W1, W2, M1(赤とする)，M2（白とする）のいずれかの6通りである。事象Aを「裏面が白」、事象Bを「表面が赤」とすると、事象Aのとき表面は、W1, W2, M1のいずれか、事象Bのとき表面はR1, R2, M1のいずれかだから、事象A∩Bのとき表面はM1のみなので

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}$$

独立

- ２つの事象A、Bについて

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

のとき、これらは独立であるという。

- これは条件付き確率を用いれば

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

ともかける。すなわち、事象Bが、事象Aの確率に何ら影響を与えないことを表す。

ベイズの定理

- 事象Aと、互いに排反（同時には起こらない）な事象B_i (i=1,…N) に対して以下が成り立つ

$$
\begin{aligned}
P\left(B_i|A\right) &= \frac{P\left(A\cap B_i\right)}{P(A)} \\
&= \frac{P\left(B_i\right)P\left(A|B_i\right)}{P\left(A\cap B_1\right)+P\left(A\cap B_2\right)+\cdots+P\left(A\cap B_N\right)} \\
&= \frac{P\left(B_i\right)P\left(A|B_i\right)}{\sum_{n=1}^{N}P\left(B_n\right)P\left(A|B_n\right)}
\end{aligned}
$$

  - B_iが原因、Aが結果である場合、$P(A|B_i)$は原因から結果への順方向の確率、$P(B_i|A)$は結果から原因への逆方向の確率を表す。工学的には結果から原因を推定するのに広く用いられる。

ベイズの定理の例題

- Aさんが住んでいる地域の天気は、70%の確率で晴れ、30%の確率で雨である。Aさんは、雨の日には必ず傘をもつが、晴れの日にも(気まぐれに)20%の確率で傘を持ってでかける。ある日、Aさんが傘をもって出かけたのを見た。その日の天気が雨である確率を求めよ。

$$P(雨|\ 傘) = \frac{P(雨)P(傘|\ 雨)}{P(晴)P(傘|\ 晴)+P(雨)P(傘|\ 雨)}$$

$$= \frac{0.3 \times 1}{0.7 \times 0.2 + 0.3 \times 1} = 0.6818\ldots$$

思ったほど高くない？
もともとのハレル確率にも
関係することに注意

ベイズの定理の適用例

- 音声認識
  - 単純化すると以下の確率を最大化する i を求める問題

$$P\left(\underline{B_i}|\underline{A}\right)=\frac{P\left(B_i\right)P\left(A|B_i\right)}{\boxed{\sum_{n=1}^{N}P\left(B_n\right)P\left(A|B_n\right)}}\propto\boxed{P\left(B_i\right)}\boxed{P\left(A|B_i\right)}$$

事象A:音声信号Aが
観測された

=P(A)であり i によら
ないので、認識には
関係ない

音響モデル:単語iを
話した時の音声信号
がAになる確率

事象B_i:単語i を話した

言語モデル:単語i を
話す確率(どんな単語
が現れやすいか)

実際には音素単位、単語単位
など階層的に行われる

確率変数

- 変数の値が確率的に定まる変数を確率変数と呼ぶ
- 離散的な確率変数
  - 例1：サイコロの目（X={1,2,3,4,5,6}）
  - 例2：文字認識における文字（X={‘a’,’b’,’c’,…}）
- 連続的な確率変数
  - 例1：電圧の測定値（X=3.4588…[V], -1.2844…[V],…)
  - 確率変数Xが連続値をとる場合、Xが正確にある値をとる確率は考えようがない（0としかいえない）。
  - X=1.000…[V]である確率は？0？
  - どんな値を考えても確率は0にしかならない？
  - 数学的には「測度 (measure)」という概念で議論される（がここでは深入りしない）

累積分布関数と確率密度関数

- 連続的な確率変数は常に、ある範囲の中にはいる確率を考える。
  - 確率変数Xが、ある値 x 以下である確率を、累積分布関数(または単に分布関数)と呼ぶ。

$$F(x) = P(X \leq x)$$

  - 累積分布関数の微分を確率密度関数と呼ぶ。

$$p(x) = \frac{dF(x)}{dx}$$

p(x)は確率密度であり、確率ではないので、1より大きい値をとることもある

  - 確率変数Xがある範囲[a, b]にはいる確率は、確率密度関数の積分により与えられる

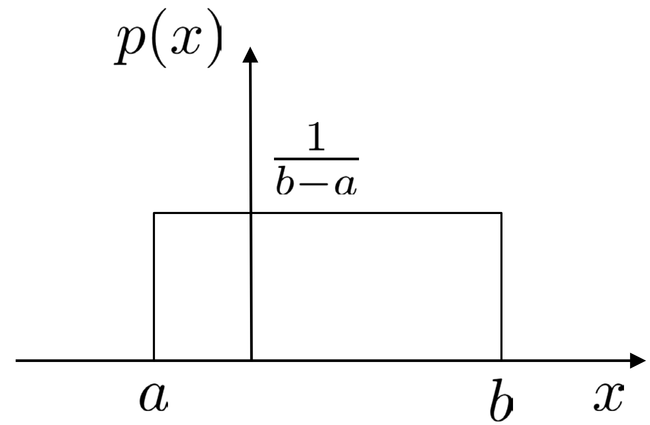
$$P(a \leq X \leq b) = \int_a^b p(x)dx$$

  - 確率密度関数は全定義域で積分した場合、値は1となる

確率密度関数の例

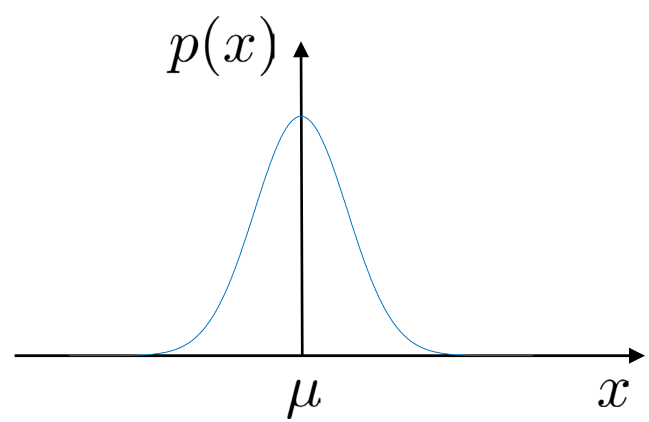
- 一様分布 (uniform distribution)

$$p(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (x < a,\ b < x) \end{cases}$$



- 正規分布 (normal distribution)
  - ガウス分布 (Gaussian distribution)とも呼ばれる
  - 計測誤差などのよいモデル（後述）

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



MATLABによる乱数の生成

- rand (m, n)
  - 各要素が、0から1までの一様分布に従う m x n サイズの行列を生成

```
>> rand(2,3)
  0.6231    0.1110    0.5463
  0.1047    0.1964    0.1038
```

ひとつひとつの要素が0から1までの一様分布から生成されている

- randn (m, n) : (最後の nはおそらく normal distributioの n)
  - 各要素が平均0、分散1の正規分布（標準正規分布）に従う m x nサイズの行列を生成

```
>> randn(2,3)
 -0.1922    1.5301   -1.0642
 -0.2741   -0.2490    1.6035
```

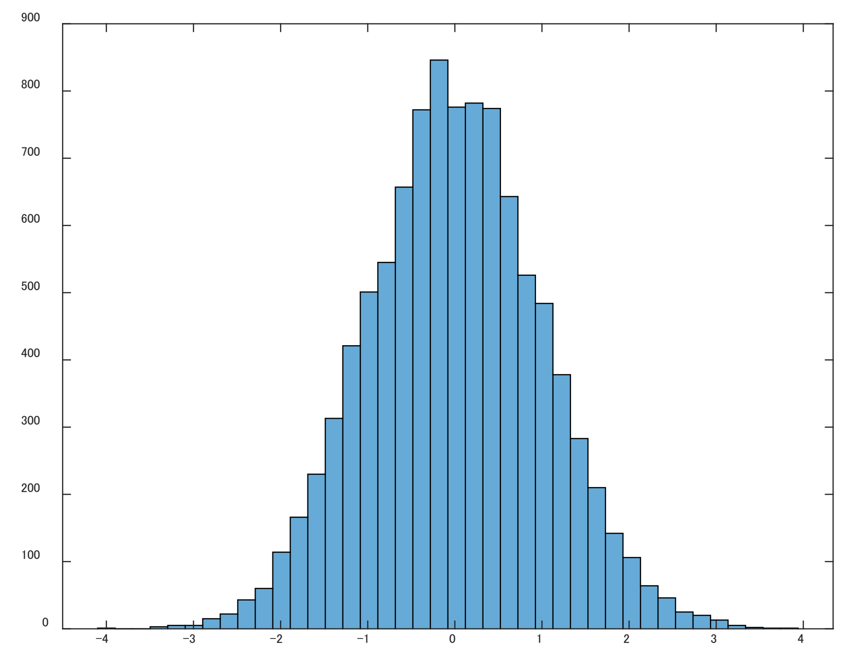
ひとつひとつの要素が平均0、分散1の正規分布から生成されている

MATLABによるヒストグラムのプロット

- histogram(X, n)
  - ベクトル（行列）Xをn個のbin (箱や容器を表す英単語、ここではヒストグラムをとる際の区間の数）

>> X=randn(1,10000);

>> histogram(X,40);



期待値

- 連続な確率変数Xに対するf(X)の期待値
(f(X)は任意の関数)

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

- 様々な統計量（分布を代表する値）が期待値として定義される
  - 例）

平均 $$\mu = E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

分散 $$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

標本平均による期待値の推定

- 十分大きいNに対して、Xの値をN回の試行得た時、近似的に以下が成り立つ

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx \simeq \frac{1}{N}\sum_{n=1}^{N} f(X_n)$$

  - これを積分の近似として行うものをモンテカルロ積分と呼ぶ

- 例題

```
>> X=randn(1,10000);
>> mean(X.^4);
```

$$\simeq \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 3$$

中心極限定理 (1/2)

- 確率変数X_nが平均μ、分散σ^2に従うとき、X_nの平均:

$$\frac{1}{N}\sum_{n=1}^{N} X_n$$

（これも確率変数）は、Nが大きくなるにつれて、
平均μ、分散σ^2/Nの正規分布に近づく。

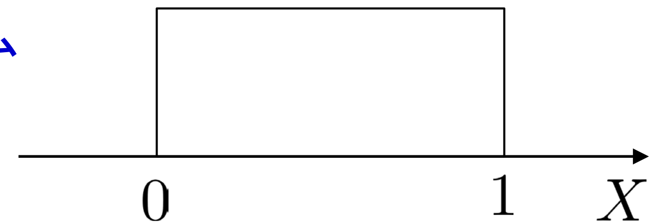
  - X_nの分布によらず、同じ確率分布に従う確率変数の平均が、標本数を大きくするにつれ、正規分布に近づいていくことを示している。
  - 正規分布の重要性の一つの理由となっている。

MATLAB練習16:中心極限定理の数値実験(1/6)

- 「0から1までの一様乱数N個を平均する」という試行を10000回繰り返し、ヒストグラムをとってみる。
- N=1の場合(平均なし)

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,10000} \end{pmatrix}$$

ヒストグラム



一様乱数そのままのヒストグラムなので、平らに近い形になるはず

MATLAB練習16:中心極限定理の数値実験(2/6)

```
>> N=1;                      % まずN=1とする
>> X=rand(N,10000);          % 各要素が、0から1までの一様分布に従うN×10000の
                             % 行列を生成。各列が1回の試行に相当
>> histogram(X,40);          % Xのヒストグラム(40個のbinを使用)を生成
```

MATLAB練習16:中心極限定理の数値実験(3/6)

- 「0から1までの一様乱数N個を平均する」という試行を10000回繰り返し、ヒストグラムをとってみる。
- N=2の場合

$$\left(\boxed{\begin{matrix} X_{1,1} \\ X_{2,1} \end{matrix}} \quad \boxed{\begin{matrix} X_{1,2} \\ X_{2,2} \end{matrix}} \quad \begin{matrix} \cdots \\ \cdots \end{matrix} \quad \boxed{\begin{matrix} X_{1,10000} \\ X_{2,10000} \end{matrix}}\right)$$

平均 ↓

$$\left(Y_{1,1} \quad Y_{1,2} \quad \cdots \quad Y_{1,10000}\right)$$

ヒストグラム ⇒

平均は変わらないはずなので
0.5を中心とした分布が
予想されるが形は？

0.5 $\quad$ Y

MATLAB練習16:中心極限定理の数値実験(4/6)

```
>> N=2;                      % N=2とする
>> X=rand(N,10000);          % 各要素が、0から1までの一様分布に従うN×10000の
                             % 行列を生成。各列が1回の試行に相当
>> Y=mean(X);                % Xの列ごとの平均を求める
>> histogram(Y,40);          % Yのヒストグラム(40個のbinを使用)を生成
```

分布の形、平均、分散がどうなっているかも確認すること

MATLAB練習16:中心極限定理の数値実験(5/6)

- 「0から1までの一様乱数N個を平均する」という試行を10000回繰り返し、ヒストグラムをとってみる。
- N=4の場合

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,10000} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,10000} \\ X_{3,1} & X_{3,2} & \cdots & X_{3,10000} \\ X_{4,1} & X_{4,2} & \cdots & X_{4,10000} \end{pmatrix}$$

平均

$$\begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,10000} \end{pmatrix}$$

ヒストグラム

平均は変わらないはずなので
0.5を中心とした分布が
予想されるが形は？

0.5 $\quad Y$

MATLAB練習16:中心極限定理の数値実験(6/6)

```
>> N=4;                    % N=4とする
>> X=rand(N,10000);        % 各要素が、0から1までの一様分布に従うN×10000の
                           % 行列を生成。各列が1回の試行に相当
>> Y=mean(X);              % Xの列ごとの平均を求める
>> histogram(Y,40);        % Yのヒストグラム(40個のbinを使用)を生成
```

分布の形、平均、分散がどうなっているかも確認すること。
余裕があれば、より大きいNに対してもやってみること。

まとめ

- 条件付き確率
- ベイズの定理
- 連続的な確率変数
- 正規分布
- 中心極限定理