知識工学

第七回　中間試験の解説，不確実な知識を用いた推論

コメント

http://papapac.com/post.php?room=知識情報工学citns

コメントを投稿したい人

教えてもらった部屋名を入れてね

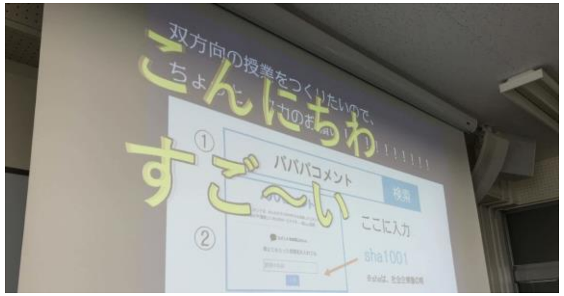
知識情報工学citns

入室

知識情報工学citnsの部屋

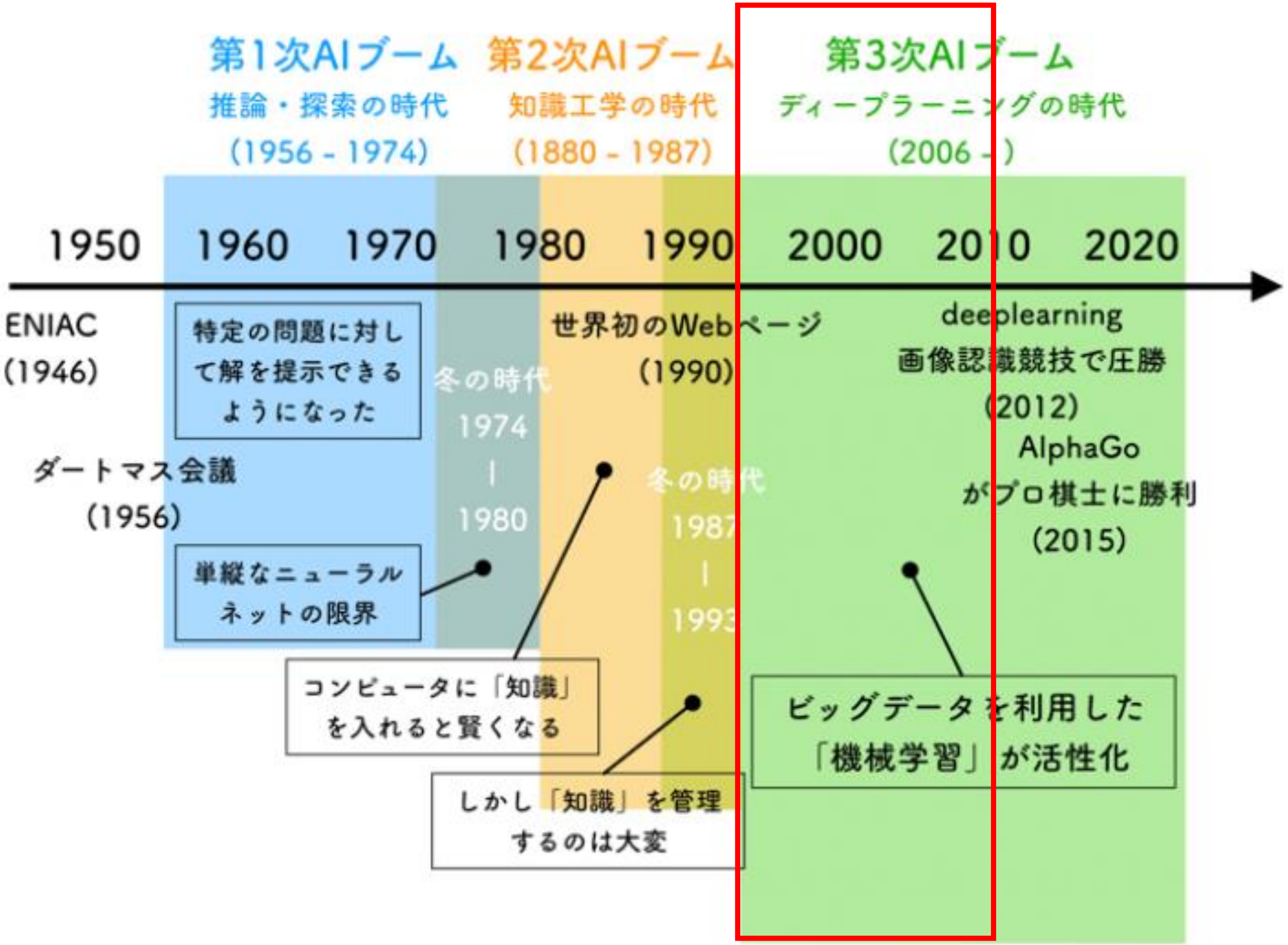
コメント送信

コメントは部外者にも見られる可能性があります。個人情報などは送信しないでください。



今日の内容

- 機械学習の基礎となる手法を学ぶ
- 現代でも使われている手法となる



現実の不確実性

- 現実世界では，常に真であるような事象は少ない
  - サイコロでさえ，確率が厳密に1/6となることは少ない

- そのため，不確実性を伴う推論，推定が必要である

条件付き確率

ある条件のもとで，その事柄が起こる確率

P(B|A)： Aが起こった下でBとなる確率

例）A:4面ダイスが1が出る．B:6面サイコロで6が出る．

P(A) P(B|A) ：AとBが同時に起こる確率

例） A:4面ダイスが1が出る．B:6面サイコロで6が出る．

Aを前提としたときのBの確率

$$P(\,B\mid A\,)$$

※PはProbability（確率）
を示している

条件付き確率

ある条件のもとで，その事柄が起こる確率

P(B|A)： Aが起こった下でBとなる確率

例）A:4面ダイスが1が出る．B:6面サイコロで6が出る．

P(B|A)= 1/6．既にAは実現している前提．

P(A) P(B|A) ：AとBが同時に起こる確率

例） A:4面ダイスが1が出る．B:6面サイコロで6が出る．

Aを前提としたときのBの確率

P(B | A)

※PはProbability（確率）を示している

条件付き確率

ある条件のもとで，その事柄が起こる確率

$P(B|A)$： Aが起こった下でBとなる確率

例）A:4面ダイスが1が出る．B:6面サイコロで6が出る．

$P(B|A) = 1/6$．既にAは実現している前提.

$P(A)\,P(B|A)$ ：AとBが同時に起こる確率

例） A:4面ダイスが1が出る．B:6面サイコロで6が出る．

$P(A)\,P(B|A) = 1/4 * 1/6 = 1/24$

$P(A \wedge B)$と同じである.

Aを前提としたときのBの確率

$$P(\ B\ |\ A\ )$$

※PはProbability（確率）を示している

トーマス・ベイズ
(1702-1761)

「この定理は，神の存在でも方程式で説明できる」

未来の確率は，過去の事象を考慮して決めるべきであるという内容を数式化した功績.

ベイズの定理

ベイズの定理

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

Bが起こったときにAとなる確率 P(A|B)を，

- Aが起こる確率． P(A)
- Bが起こる確率． P(B)
- Aが起こったときにBとなる確率． P(B|A)

を用いて求めることができる．

ベイズの定理を導出

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$ ※$P(A) > 0$

両辺に$P(A)$をかける

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

代入

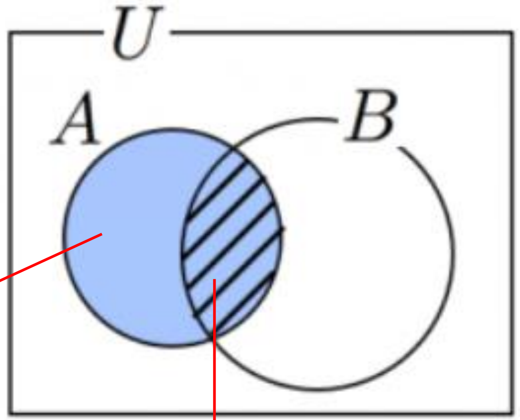
条件付き確率の公式は
AとBをひっくり返しても成立

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ ※$P(B) > 0$

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

ベイズの定理

はじめの式で，なぜP(B|A)をP(A∧B)/P(A)で表せるかは，ベン図を考えるとわかりやすい．



全体の中で，
Aが起こる確率

Aのもとで，Bが起こる確率

練習問題：結果から原因の確率を求める

- 調査から，無作為に選んだメールの 20 ％は迷惑メール，80％が一般メールであると明らかになっている．
- 迷惑メールが「令和最新」という単語を含んでいる確率は30％，一般メールが「令和最新」という単語を含んでいる確率は4％である．
- このとき，無作為に選んだメールが「令和最新」という単語を含んでいた場合，これが迷惑メールである確率は？

2分後にヒントが出ます．

練習問題：結果から原因の確率を求める

- 調査から，無作為に選んだメールの 20 %は迷惑メール，80%が一般メールであると明らかになっている.
- 迷惑メールが「令和最新」という単語を含んでいる確率は30%，一般メールが「令和最新」という単語を含んでいる確率は4%である.
- このとき，無作為に選んだメールが「令和最新」という単語を含んでいた場合，これが迷惑メールである確率は？

ヒント

A：迷惑メールである
B：「令和最新」という単語を含んでいる
として$P(A|B)$を求めればよい.

答え

A：迷惑メールである
B：「令和最新」という単語を含んでいる

① 迷惑メールである確率：20%
② 迷惑メールに「令和最新」がある確率：30%
③ 一般メールである確率：80%
④ 一般メールに「令和最新」がある確率：4%

「令和最新」があるときに迷惑メールである確率はP(A|B)

ベイズの定理

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

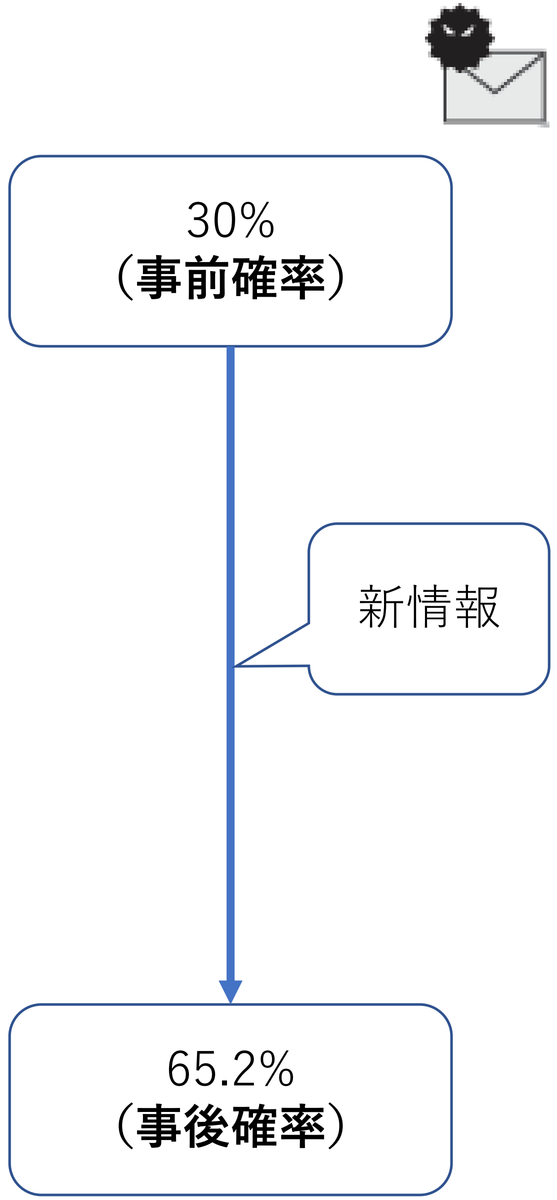
P(B)=0.2*0.3 + 0.8*0.4=0.092　(①〜④より)
P(B|A)=0.3（②より）
P(A)=0.2（①より）
P(A|B)=（0.3*0.2）/ 0.092 ≒ 0.652

答え　65.2%

ベイズ更新

- 迷惑メールの確率は本来30%だが，「令和最新」が含まれているという情報により，迷惑メールである確率が65.2%になった．
- 新情報により確率が変化したと見なせる

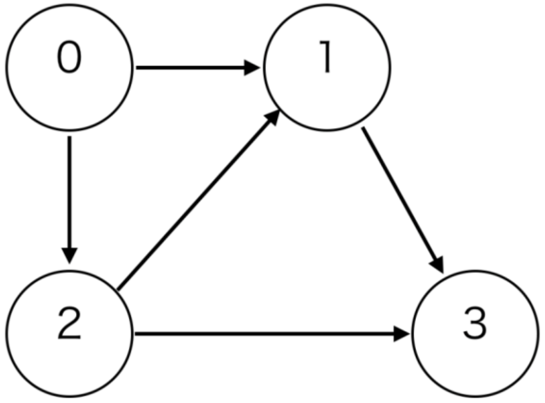
- はじめに与えられた確率が，情報によって更新されることを**ベイズ更新**という．



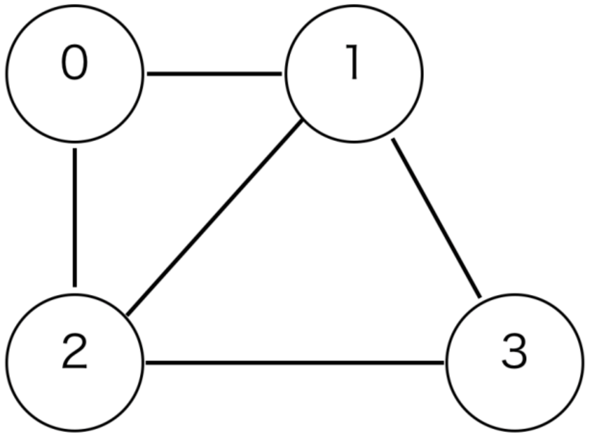
ベイジアンネットワーク

- 物事の因果関係を矢印でつないだ有向グラフ
- 丸の部分を**ノード**，矢印を**リンク**という

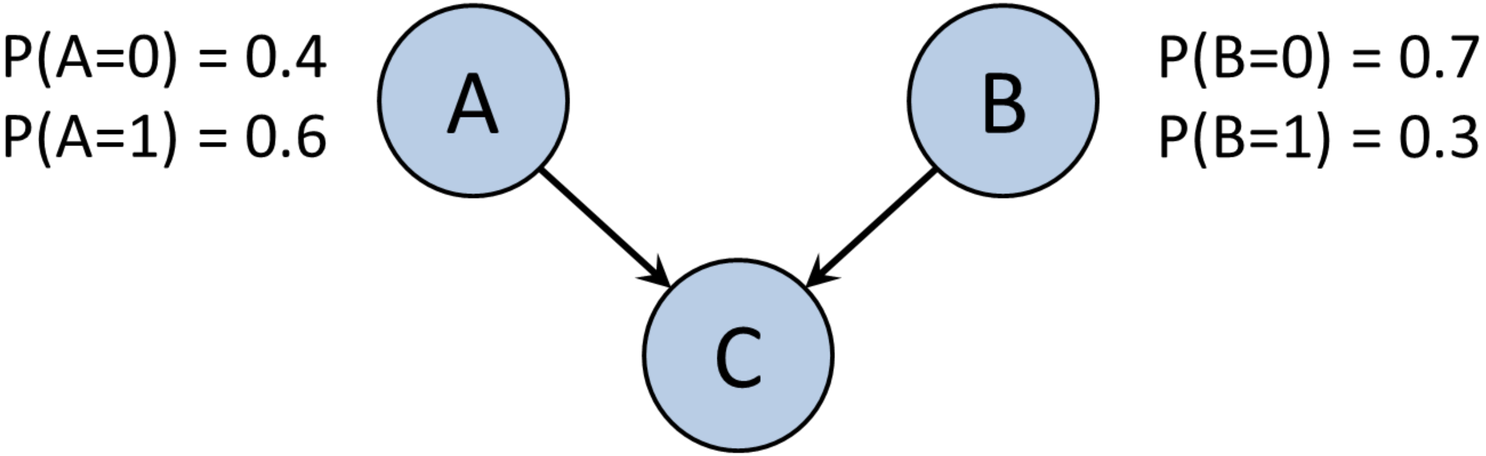
有向グラフ



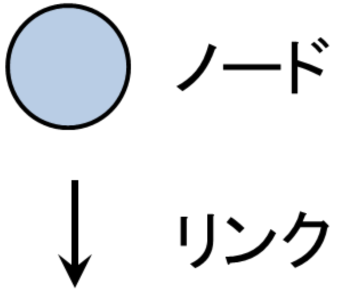
無向グラフ



ベイジアンネットワークの例



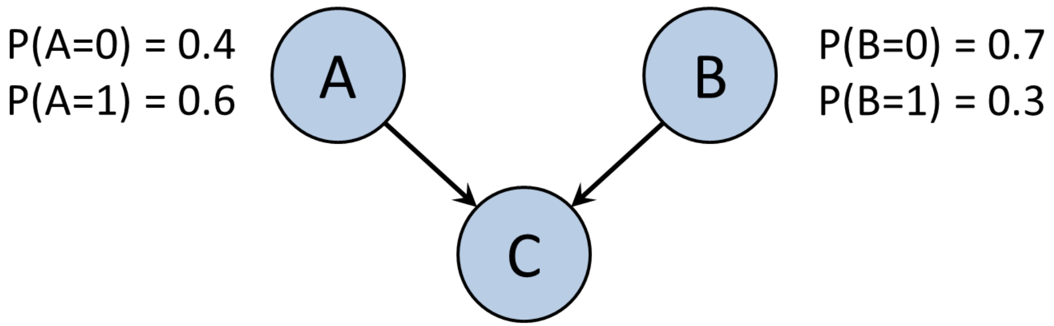
A	B	P(C|A,B)	
		C=0	C=1
1	1	0.3	0.7
1	0	0.5	0.5
0	1	0.2	0.8
0	0	0.6	0.4



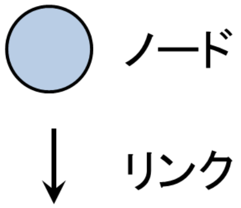
P(C｜A,B)は，AとBが同時に起こったときにCとなる確率を示す．P(C｜A∧B)と同じである．

事後確率の計算

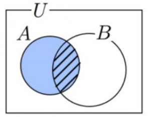
人につられやすいCさんが怒っている(C=1)とき，Aさんが怒っている（A=1）確率を求める．



A	B	P(C|A,B)	
		C=0	C=1
1	1	0.3	0.7
1	0	0.5	0.5
0	1	0.2	0.8
0	0	0.6	0.4



$P(A=1|C=1)$は，$P(A=1, C=1) \div P(C=1)$．



$P(C=1)$は，その状態になる時の確率の和で求められる．

$P(C=1|A=1,B=1)P(A=1)P(B=1) = 0.7 \times 0.6 \times 0.3 = 0.126$ … ①
$P(C=1|A=1,B=0)P(A=1)P(B=0) = 0.5 \times 0.6 \times 0.7 = 0.21$ … ②
$P(C=1|A=0,B=1)P(A=0)P(B=1) = 0.8 \times 0.4 \times 0.3 = 0.096$ … ③
$P(C=1|A=0,B=0)P(A=0)P(B=0) = 0.4 \times 0.4 \times 0.7 = 0.112$ … ④
$P(C=1)=0.126+0.21+0.096+0.112 =0.544$

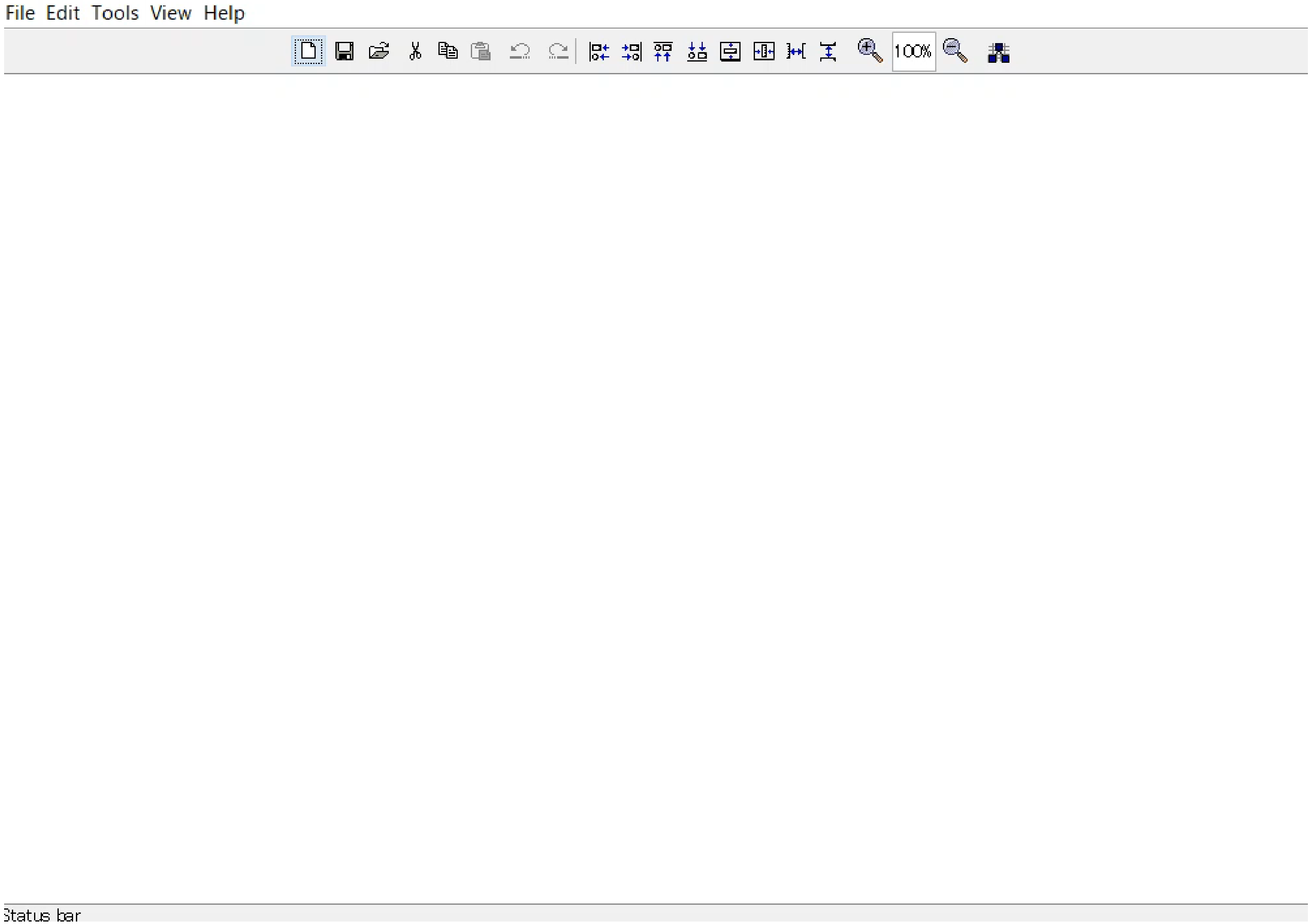
$P(A=1, C=1)$も，その状態になる確率の和．

$P(A=1, C=1)$ = ① + ② = $0.126 + 0.21 = 0.336$

従って，

$P(A=1|C=1) = P(A=1, C=1) \div P(C=1) = 0.336/0.544=0.6176$

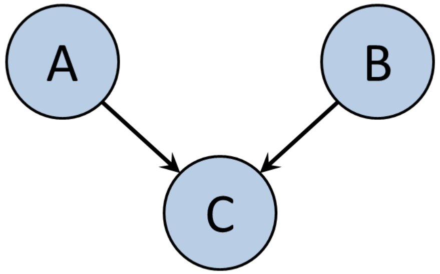
ツールでの検証



$P(A=1|C=1) = 0.6176$

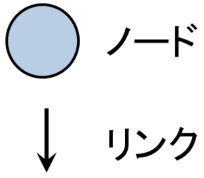
Value 1＝ 0 の状態
Value 2＝ 1 の状態

P(A=0) = 0.4
P(A=1) = 0.6



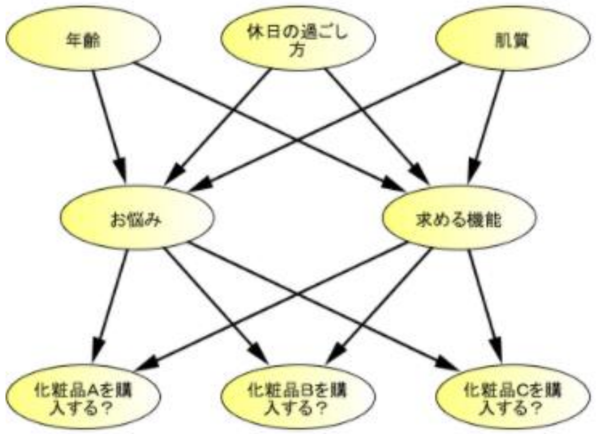
P(B=0) = 0.7
P(B=1) = 0.3

A	B	P(C|A,B)	
		C=0	C=1
1	1	0.3	0.7
1	0	0.5	0.5
0	1	0.2	0.8
0	0	0.6	0.4



ベイジアンネットワークの活用例

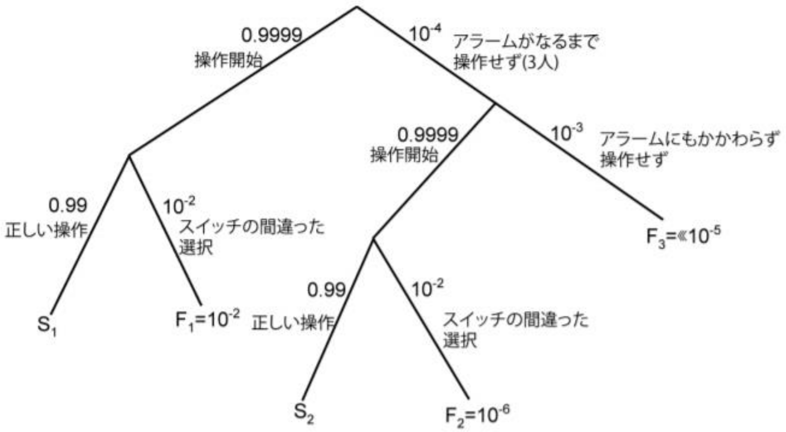
定性的なデータを得られる際に,
分析や予測で用いられることが多い

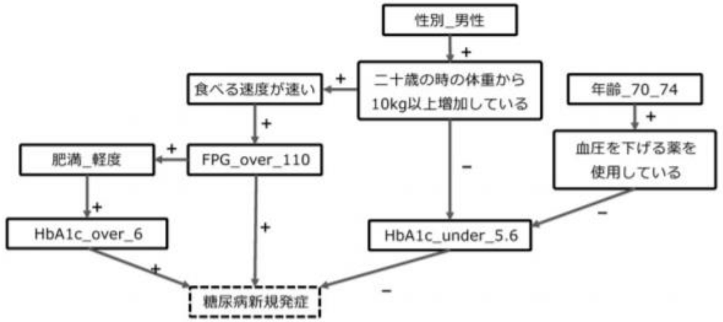


- マーケティング調査

- システムの故障診断

- 健康予測





休憩

最尤推定

- 尤度とは「もっとも（尤も）らしさの度合い」
  パラメータθを指定したとき，今手持ちのデータが再現できる確率
- 尤度がもっとも高くなるように，　θを推定

例）コインを2回投げ，はじめは表，次は裏が出た．

データ：一回表，一回裏

θ　：コインが表になる確率

最尤推定

- 尤度とは「もっとも（尤も）らしさの度合い」
  パラメータθを指定したとき，今手持ちのデータが再現できる確率
- 尤度がもっとも高くなるように，　θを推定

例）コインを2回投げ，はじめは表，次は裏が出た．

データ：一回表，一回裏

θ　：コインが表になる確率

$\underbrace{1/3}_{\text{表の確率}} * \underbrace{(1 - 1/3)}_{\text{裏の確率}} = 1/3 * 2/3 = \underbrace{2/9}_{\text{今回のデータが生じる確率}}$

尤度

最尤推定

- 尤度とは「もっとも（尤も）らしさの度合い」
  パラメータ θ を指定したとき，今手持ちのデータが再現できる確率
- 尤度がもっとも高くなるように， θ を推定

例）コインを2回投げ，はじめは表，次は裏が出た．

データ：一回表，一回裏

θ ：コインが表になる確率

$$\underbrace{1/2}_{\text{表の確率}} * \underbrace{(1 - 1/2)}_{\text{裏の確率}} = 1/2 * 1/2 = \underbrace{1/4}_{\text{今回のデータが生じる確率}}$$

尤度

尤度の求め方

- 観測データがx_1,x_2…x_nという値で出現した場合に，尤度は以下のようにして求めることが出来る

$$尤度 = P(x_1)P(x_2).... P(x_n)$$

- 見覚えある計算式ではないですか？？

尤度と同時確率の関係

求め方は同じ式であるが，考え方が逆転している

<u>同時確率</u>

ある事象が独立して同時に起こる確率を予想する考え方

<u>尤度</u>

データが出揃ったとして，その確率からパラメータを推定する.

それらのデータに対して，あるパラメータの確率分布を当てはめた時，どれだけ尤もらしいかを計算する.

最尤推定の特徴

- 得られたデータのみで考える
- サンプルサイズが大きくなるほど，尤度は小さくなる．
  - 例）100回中60回表がでる確率L（尤度関数）．

$$L(\theta) = {}_{100}\mathrm{C}_{60}\theta^{60}(1-\theta)^{40}$$

- このままでは扱いづらいため，通常は対数をとる．

最大尤度の求め方

log L(θ) が最大値を取る

⬇

log L(θ) が極大値を取る

⬇

log L(θ) を θ で微分したものが0になる

$$(\log x)' = \frac{1}{x}$$

$$\begin{aligned}\log L(\theta) &= \log\left({}_{100}\mathrm{C}_{60}\theta^{60}(1-\theta)^{40}\right)\\ &= \log\left({}_{100}\mathrm{C}_{60}\right) + \log\left(\theta^{60}\right) + \log\left((1-\theta)^{40}\right)\\ &= \log\left({}_{100}\mathrm{C}_{60}\right) + 60\log(\theta) + 40\log(1-\theta)\end{aligned}$$

$$\frac{d}{d\theta}\log L(\theta) = \frac{60}{\theta} - \frac{40}{1-\theta} = 0$$

$$60(1-\theta) - 40\theta = 0$$

$$\theta = 0.6$$

例題

コインが1枚あり，表が出る確率をθとする．このコインをn回投げたところ，x回表が出た．このときのθを最尤推定で求めよ．

例題

コインが1枚あり，表が出る確率を θ とする．このコインをn回投げたところ，x回表が出た．このときの θ を最尤推定で求めよ．

$$L(\theta) = {}_nC_x \theta^x (1-\theta)^{n-x}$$

$$\begin{aligned}
\log L(\theta) &= log[{}_nC_x \theta^x (1-\theta)^{n-x}] \\
&= log[\tfrac{n!}{x!(n-x)!}\theta^x (1-\theta)^{n-x}] \\
&= log(n!) - log(x!) - log(n-x)! + log\theta^x + log(1-\theta)^{n-x} \\
&= log(n!) - log(x!) - log(n-x)! + xlog\theta + (n-x)log(1-\theta)
\end{aligned}$$

例題

コインが1枚あり，表が出る確率をθとする．このコインをn回投げたところ，x回表が出た．このときのθを最尤推定で求めよ．

logL(θ) $= log(n!) - log(x!) - log(n-x)! + xlog\theta + (n-x)log(1-\theta)$

θで微分し，

logL'(θ) = $\frac{x}{\theta} - \frac{n-x}{1-\theta} = \frac{x(1-\theta)-(n-x)\theta}{\theta(1-\theta)} = \frac{x-n\theta}{\theta(1-\theta)}$

これが0のとき最大になるので，$x - n\theta = 0$ より，$\theta = \frac{x}{n}$

パラメータの一般化された求め方

$$\hat{\theta} = \underset{\theta}{\arg\max}\ p(y_o|\theta)$$

観測されたデータが出る確率を最大にするよう，パラメータを最適化

- y_o:観測データ
- θ :モデルのパラメータ（表の確率・平均値など）

- モデル：前提となる分布．二項分布なら ${}_nC_k\, p^k(1-p)^{n-k}$ で確率が出る

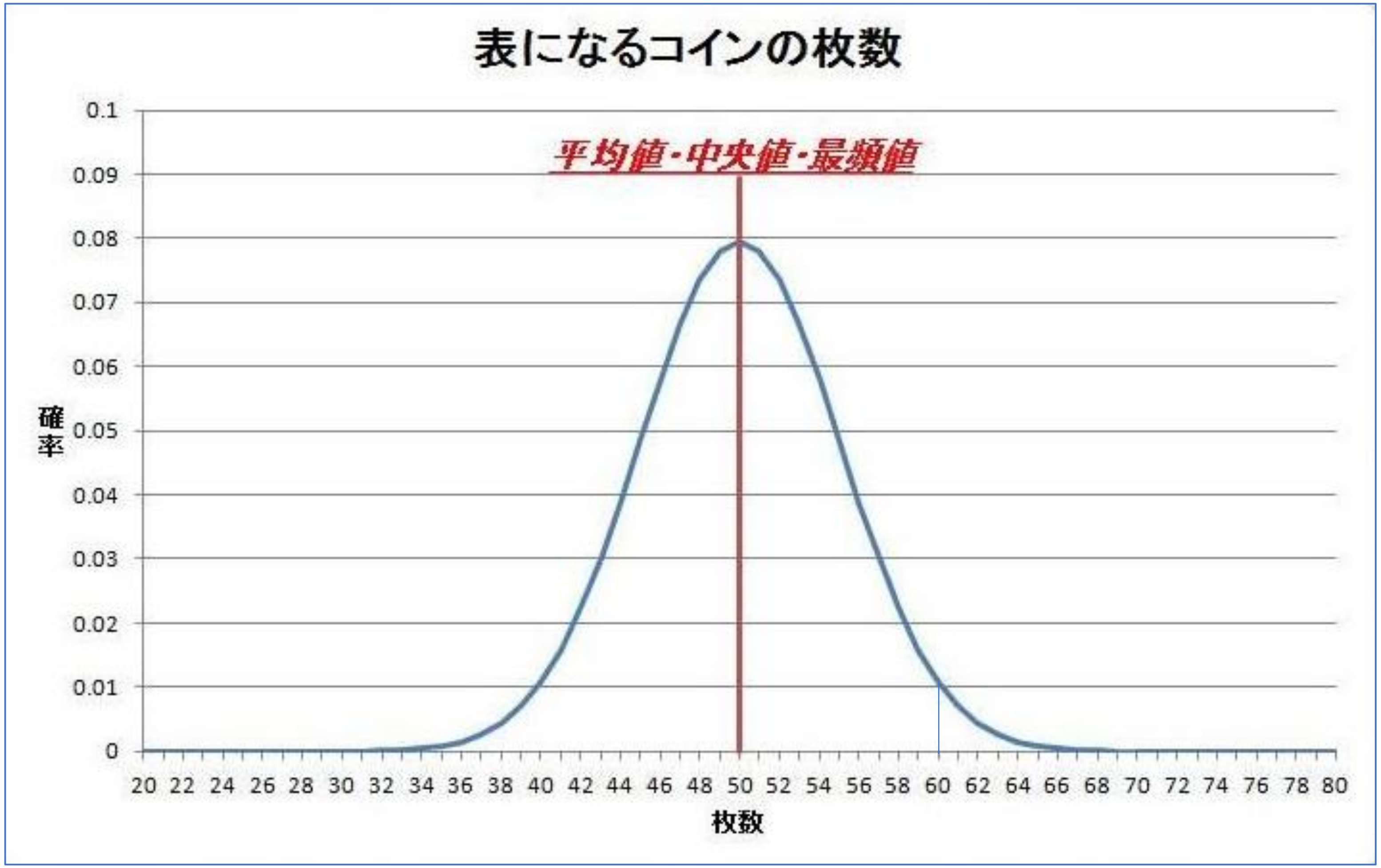
モデル（確率密度分布）

右図は正規分布と呼ばれるモデル.
身長などがこの分布になる.

高さは確率を表す.
全体の面積を１として考える.

確率を面積で表すことも可能.

例）表になるコインが60枚以上となる確率は，60のところから右側の面積を求めれば良い.



正規分布の特徴

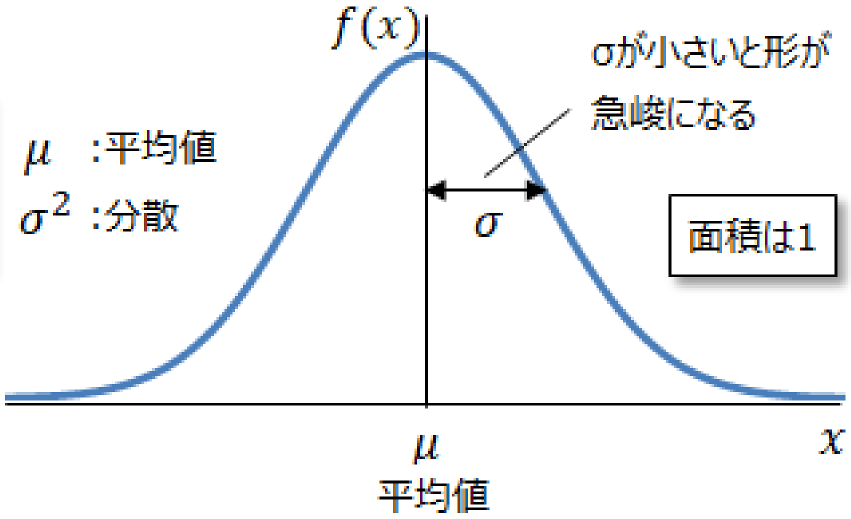
- 一番山の高いところが平均値（μ）を示す.
- 標準偏差（σ）が大きいほど広がりが大きくなる.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

標準偏差 $\sigma = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(x-\mu)^2}$

μ :平均値

σ^2 :分散



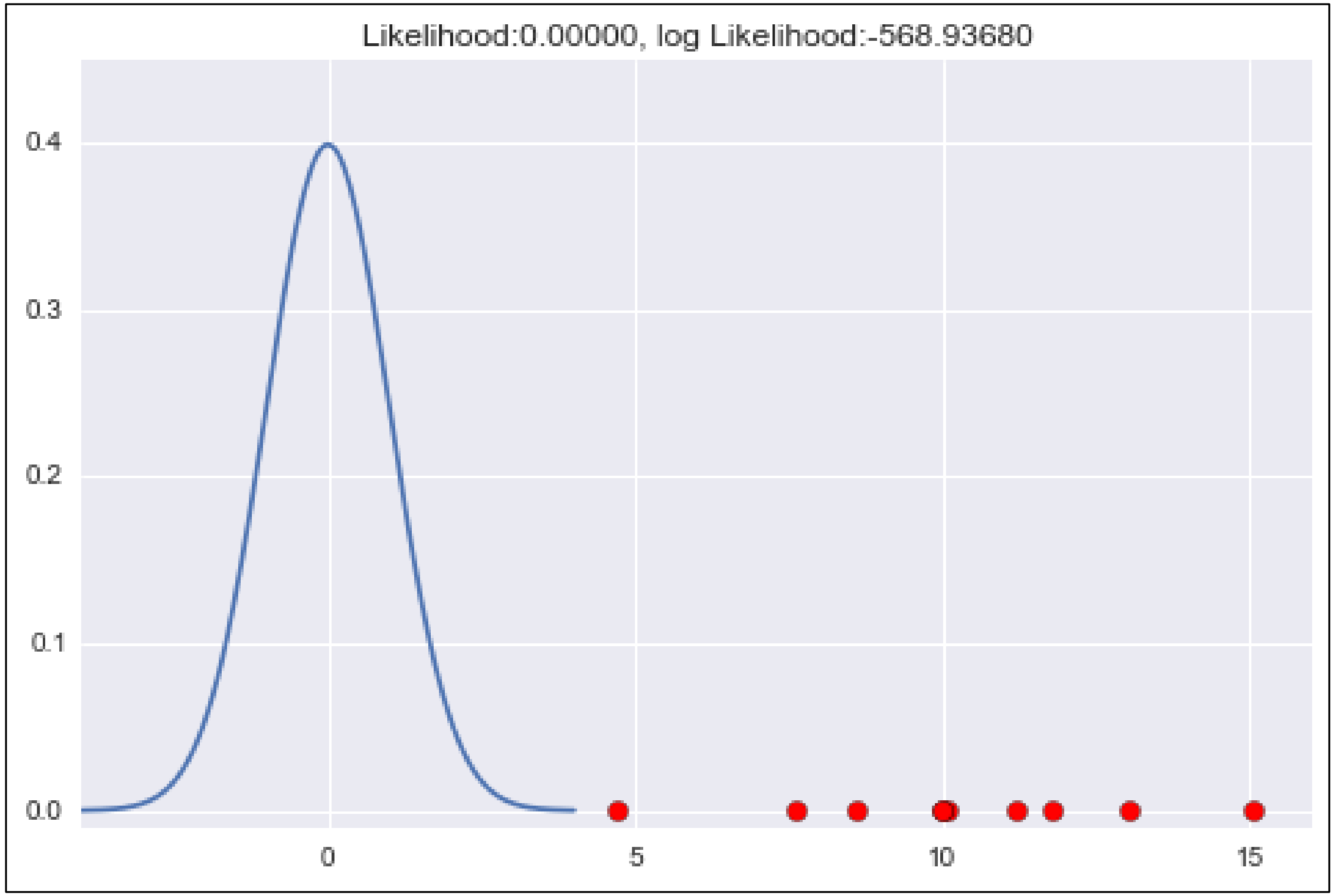
初期状態

<u>現在のパラメータ</u>

$\mu=0,\ \sigma=1$

- 丸い点がデータ
- 1つも分布内にない
  ＝尤度が 0

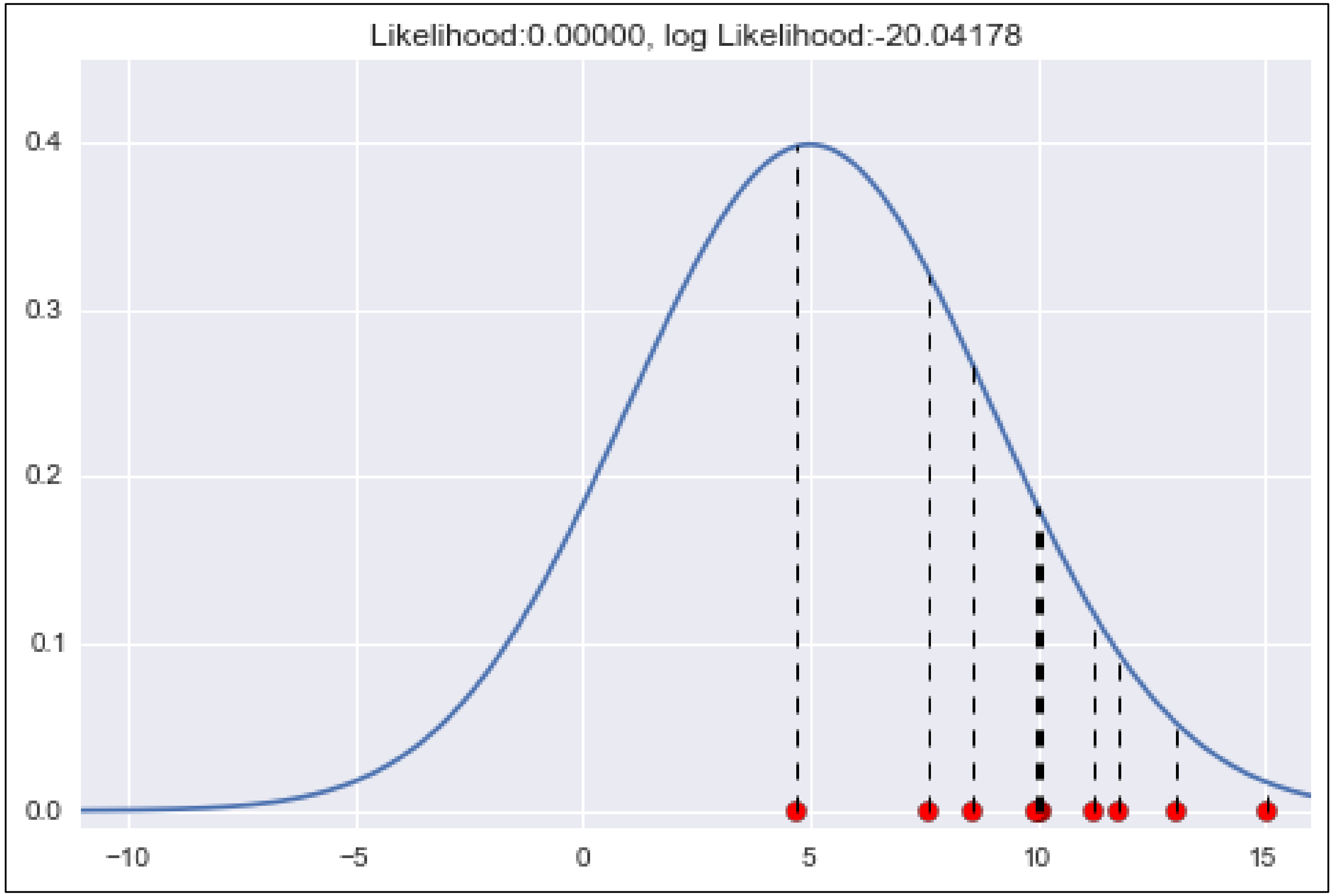
これは，0が出る確率が40%（高さ），4以降が出る確率は0%の分布



パラメータを動かす

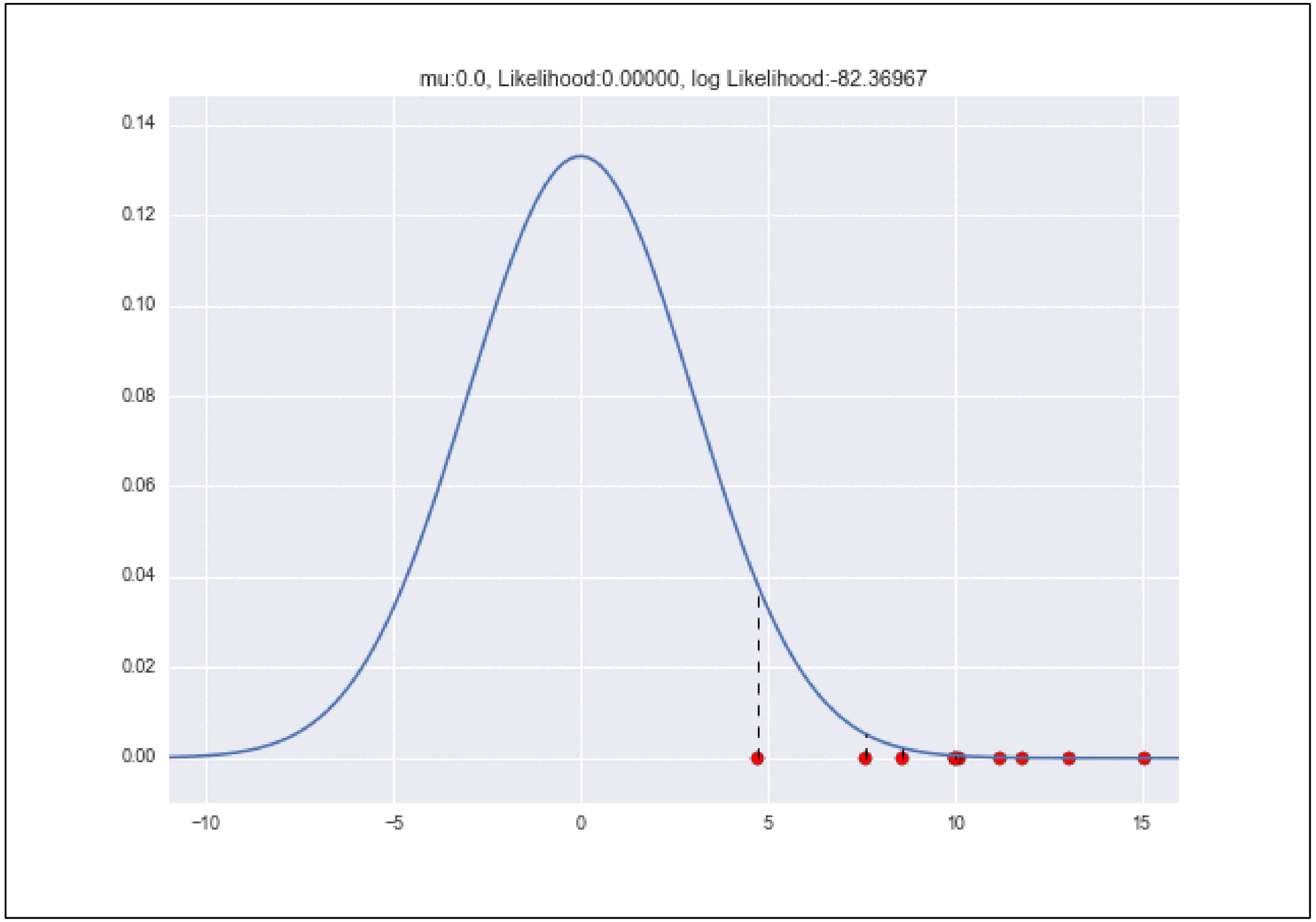
- μ =5
- σ =4

- 縦の線は確率なので，尤度が少しだけ大きくなった



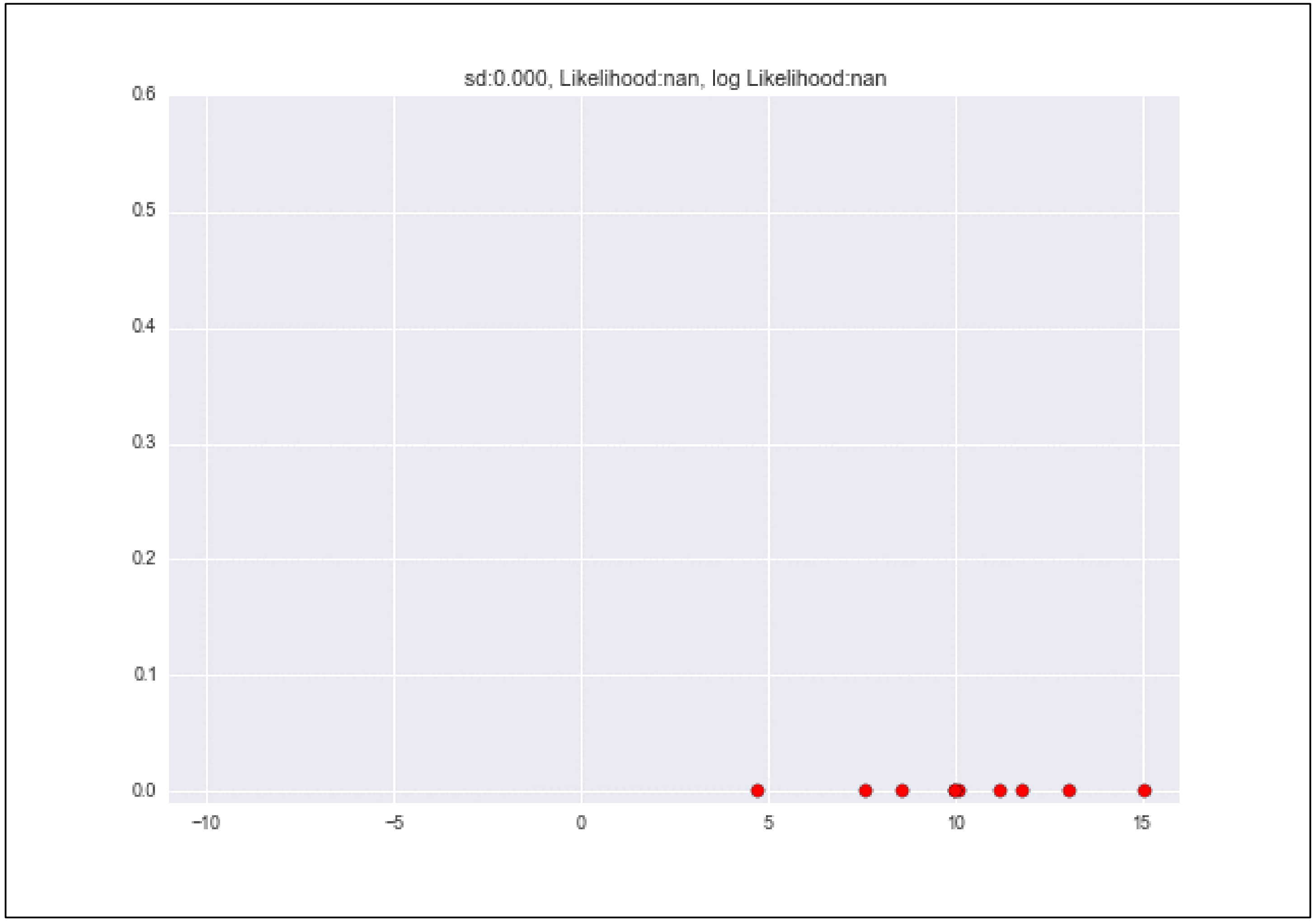
平均を変化させる

- μ =10 で尤度（対数）が最大になっている

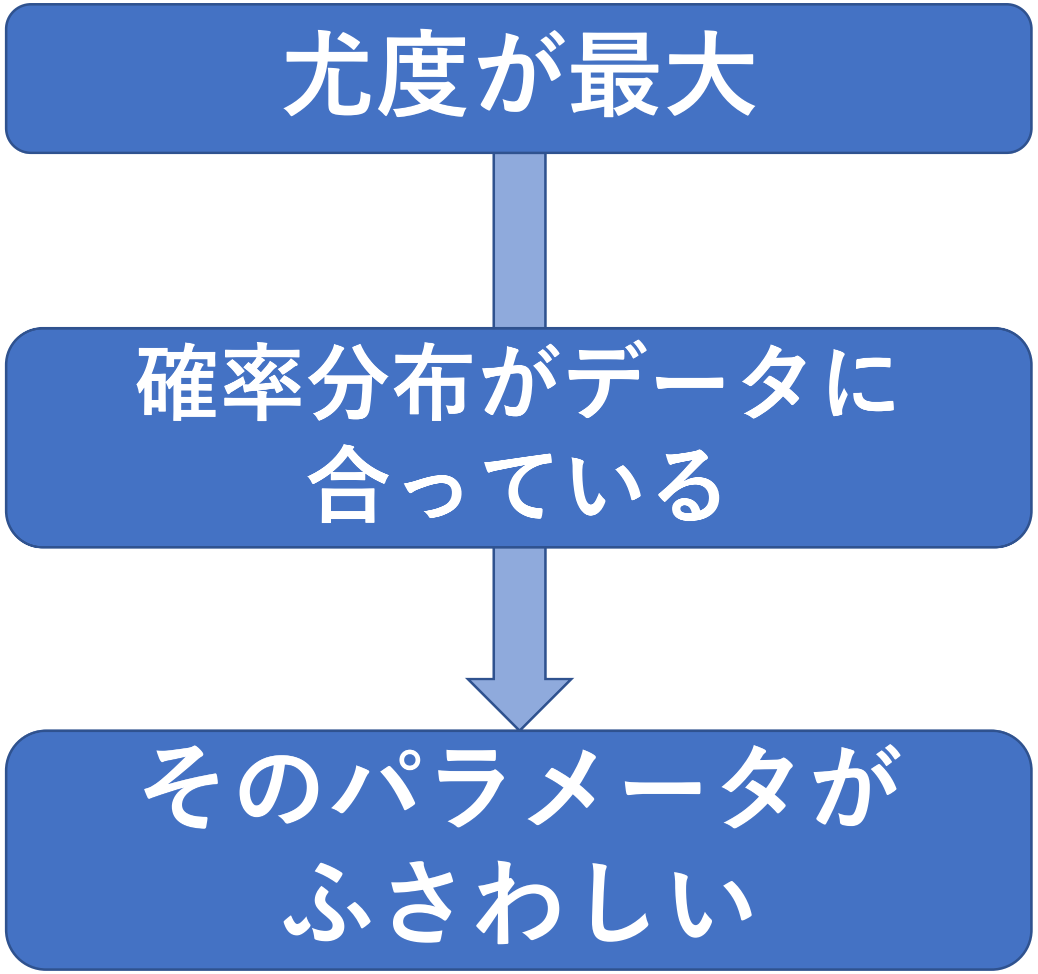


標準偏差を変化させる

- σ =2.7くらいで
  尤度が最大になる

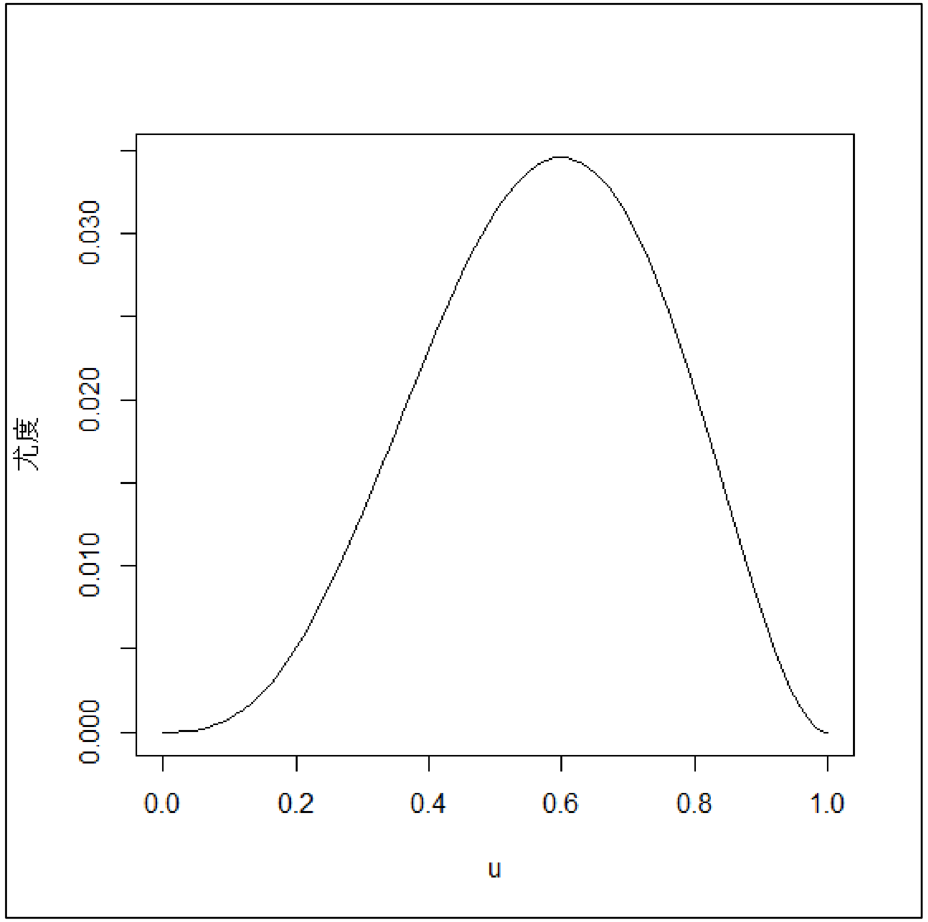


尤度が最大ということは



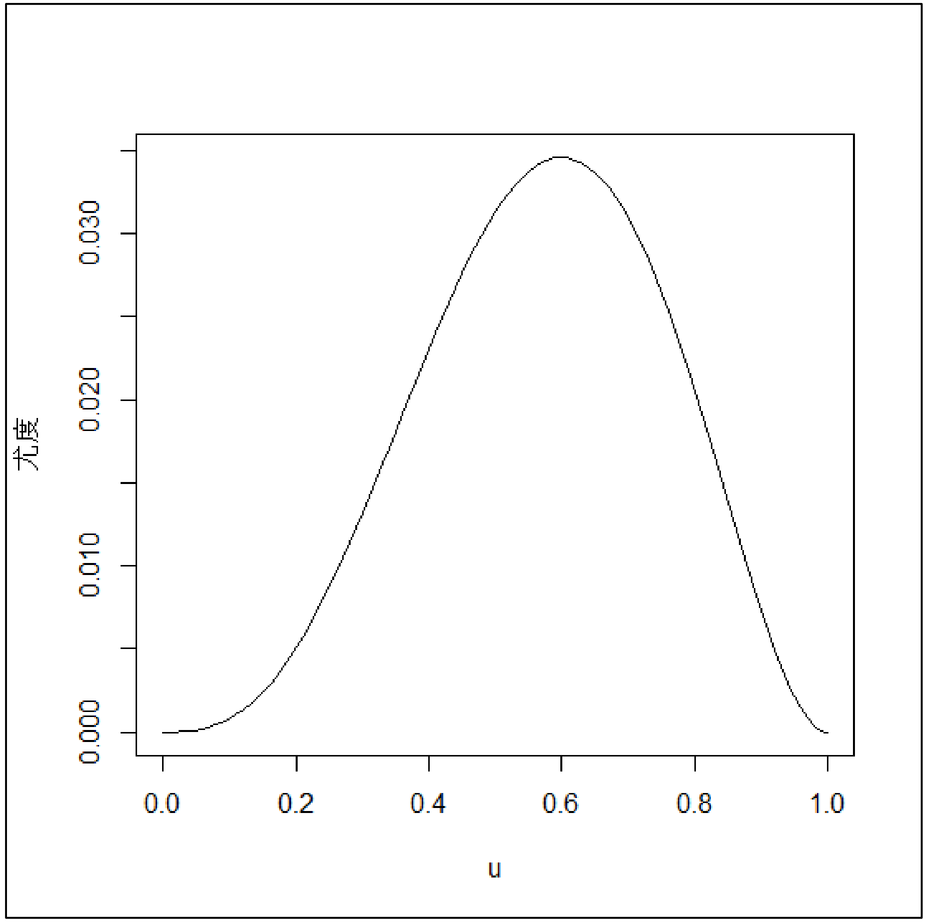
最尤推定法のRプログラム

```
coin <- c(1, 0, 1, 1, 0)

#尤度関数を作成.
Likelihood <- function(x){
  return(function(u){
    L <- 1
    for(xn in x){
      L <- L * u^(xn) * (1 - u)^(1-xn)
    }
    return(L)
  })
}

(Likelihood(coin))(0.5)
(Likelihood(coin))(0.1)

plot(Likelihood(coin), 0, 1, xlab="u", ylab="尤度")
```



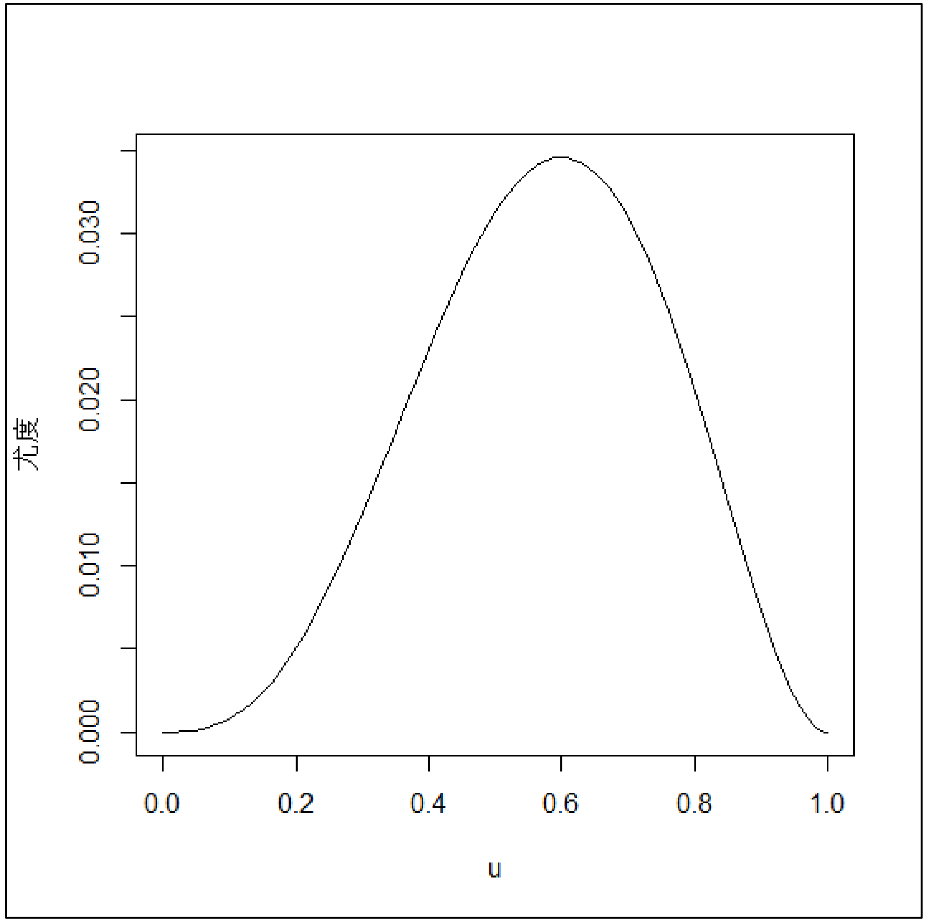
最尤推定法のRプログラム

```
# 最大となる尤度を求める
optimize(f=Likelihood(coin), c(0, 1), maximum=TRUE)

$maximum
[1] 0.6000006

$objective
[1] 0.03456
```



最尤推定法のRプログラム

```
# 尤度を対数で計算する
LogLikelihood <- function(x){
  return(function(u){
    L <- 1
    for(xn in x){
      L <- L * u^(xn) * (1 - u)^(1-xn)
    }
    return(log(L))
  })
}
plot(LogLikelihood(coin), 0, 1, xlab="u", ylab="対数尤度")

>optimize(f=LogLikelihood(coin), c(0, 1), maximum=TRUE)
$maximum
[1] 0.5999985

$objective
[1] -3.365058
```



最尤推定法のRプログラム

平均と標準偏差などの2パラメータの場合.

1パラメータの時の正規分布における
確率密度関数

$$p(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

この時の尤度関数

$$\prod_{i=1}^{n} p(x_i|\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

その対数を取って式を変形（これのμ
とσ^2を動かして最大化させたい）

$$-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

最尤推定法のRプログラム

平均と標準偏差などの2パラメータの場合.

1パラメータの時の正規分布における
確率密度関数

$$p(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

この時の尤度関数

$$\prod_{i=1}^{n} p(x_i|\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

その対数を取って式を変形（これのμ
とσ^2を動かして最大化させたい）

$$-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

最尤推定法のRプログラム

```
# 使用するデータ.
# cars$speedの平均（μ）と分散（σ^2）を推定し求めてみる
cars$speed

log_likelihood_for_norm <- function(x){
  return(function(par){
    mu <- par[1]
    sigma2 <- par[2]
    - length(x) / 2 * log(sigma2) - 1 / 2 * sum((x - mu)^2) / sigma2
  })
}

optim(par = c(5, 10), fn = log_likelihood_for_norm(cars$speed), control
= list(fnscale = -1))
```

最尤推定法のRプログラム

<u>正解値</u>

mean(cars$speed)

[1] 15.4　　　　　　：平均値

> var(cars$speed)

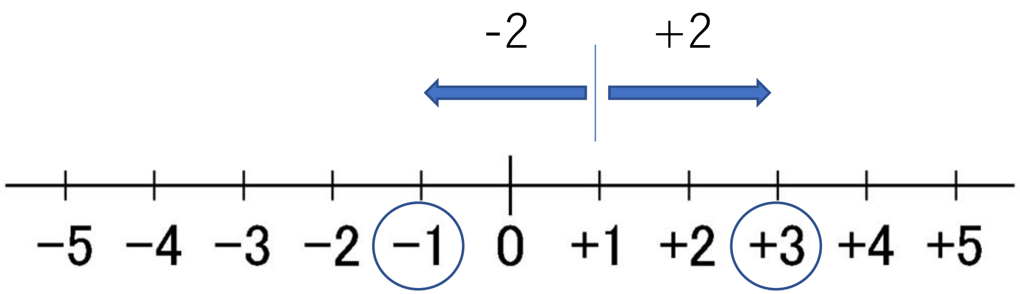
[1] 27.95918　　　：分散

```
$par
[1] 15.39812 27.40184      推定結果

$value
[1] -107.7636

$counts
function gradient
      75       NA

$convergence
[1] 0

$message
NULL
```

付録：標準偏差と分散

- データの散らばり具合を見る指標

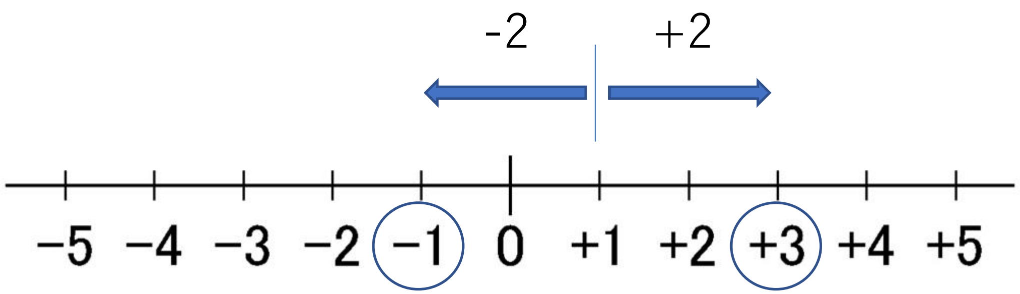
例）得られたデータが，-1と+3であった．
データの散らばり度合いは？
　→ 平均値から離れている距離の平均



平均値(μ)を出す式は　$1/n(x_1+x_2+\cdots x_n)$
距離を出す式は，　$x_n - \mu$
距離の平均を計算するとどうなる？

付録：標準偏差と分散

- データの散らばり具合を見る指標



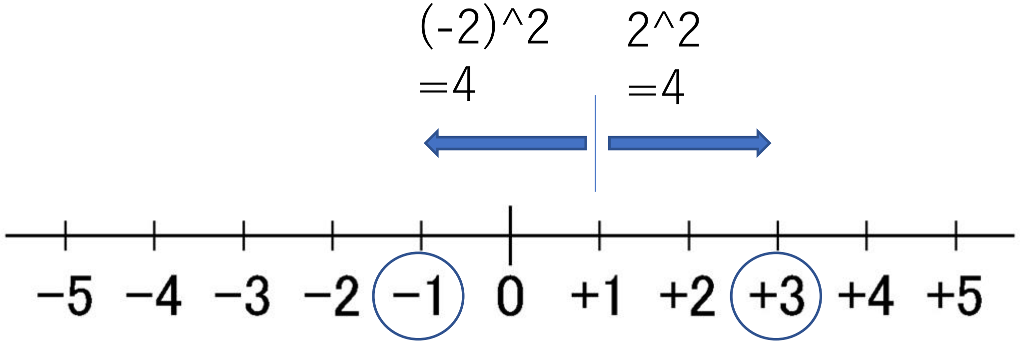
例）得られたデータが，-1と+3であった．
データの散らばり度合いは？
→ 平均値から離れている距離の平均

平均値(μ)を出す式は　$1/n(x_1+x_2+\cdots x_n)$
距離を出す式は，　$x_n - \mu$
距離の平均を計算するとどうなる？

距離の平均が0になってしまう！
つまり，散らばり度合いは0．出てきた値は全て平均値と同じ．

付録：標準偏差と分散

- データの散らばり具合を見る指標

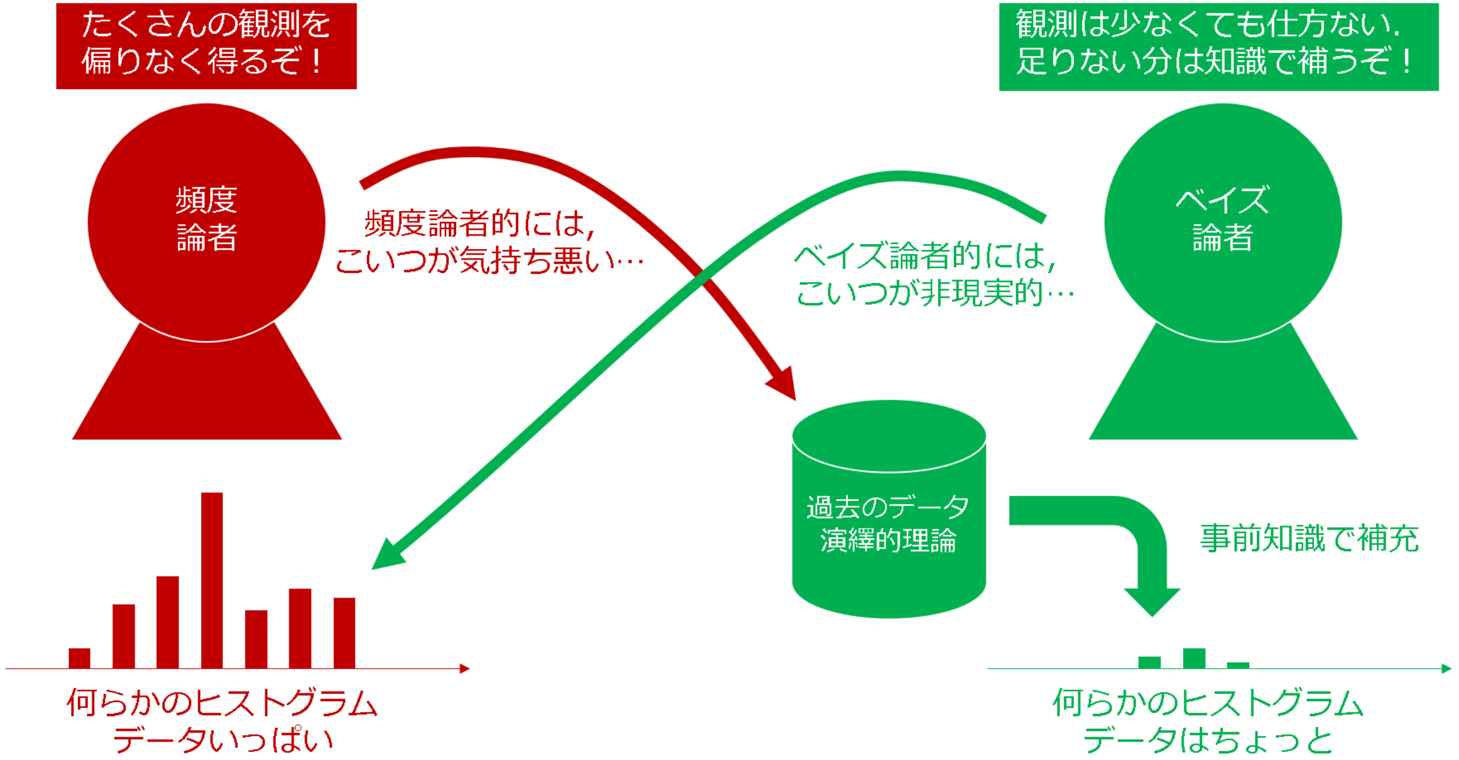


- 平均値から距離をそのまま合計すると打ち消し合ってしまう．
- 2乗すれば+になるから，距離を足していけるね！

- 距離を2乗すると，その平均は4．（**分散**）
→ けどその平均値って2乗したときのものでしょ？
→ じゃあ，平方根をとって2乗を解消し，2となりました．（**標準偏差**）

ベイズ推定との違い

- 最尤推定は，出てきたデータのみに着目する．
- ベイズ推定は，出てきたデータに加えて以前からの知識も活用する．



レビューシートの提出

- 今日の授業に関するレビューシートを，manabaから提出すること.

後日不明点があれば，多胡まで.

７号館５階　第９実験室内　第９研究室

tago@net.it-chiba.ac.jp