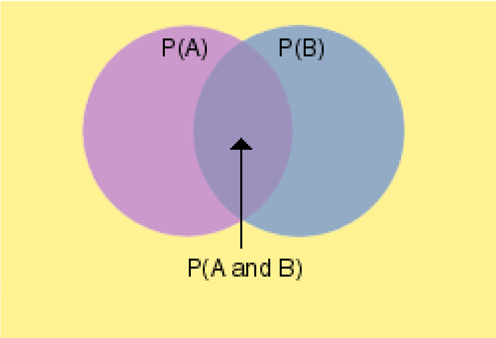
ベイジアンになろう

電子情報工学科
伊庭　斉志

確率の復習



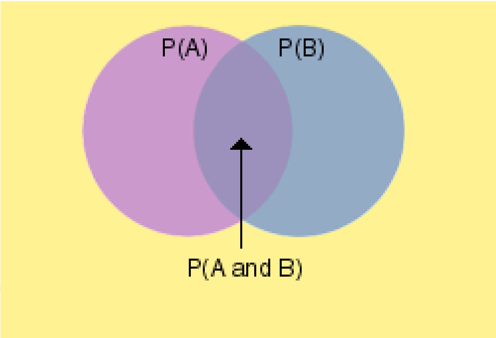
条件付き確率

Aが起こったという条件でBが起きるという事象を $B|A$ であらわす
その確率　$P(B|A)$　を条件AのもとでのBの起きる条件付き確率といい、

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

で定義する

確率の復習

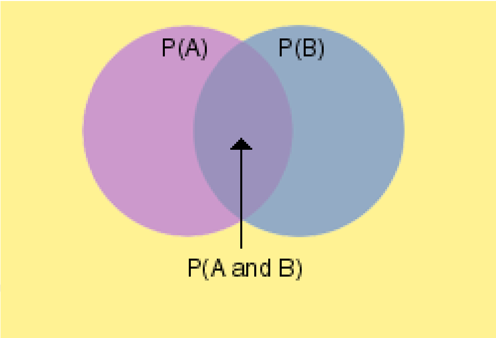


周辺確率

Aが起こったとときに同時にB1,B2,･･･Bmのいずれかが起きるとする。ただしB1,B2,･･･Bm は同時には起こらない。
このとき、Aがおこる確率は以下のように表わされる。

$$P(A) = \sum_{i=1}^{m} P(A, B_i) = \sum_{i=1}^{m} P(A \mid B_i) P(B_i)$$

確率の復習



乗法定理

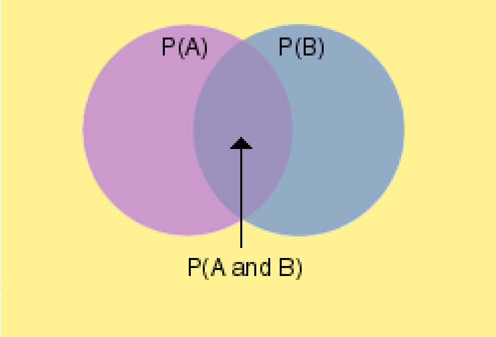
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

から

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

が得られる

例題



- あなたは乳ガンの検査(マンモグラフィ)で陽性になった。
- この検査では、乳ガンになっている人を見つける正確さは90%、なっていない人を見つける正確さは93%である。
- あなたが本当にガンである確率はいくつか？
- ただし、病気の発生確率は0.8%である。

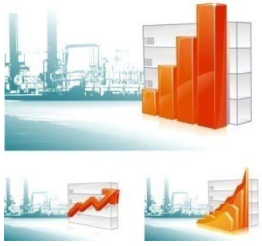
- 検査は90%近く正確なのだから、非常に可能性がある？？　(たいていの医者はこう答えた)

- 本当か？

ベイズの定理

- 用途
  結果が起きたとき、その原因を調べる
- 結果の確率（事前確率）
- 原因の確率（事後確率）

- 原因　$A_1, A_2, \cdots, A_n$（n個の排反）
- 結果　E



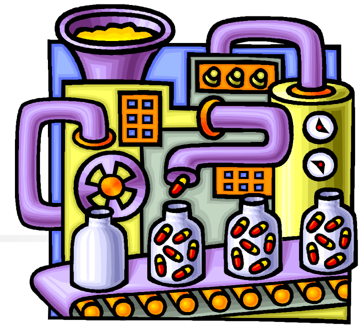
ベイズの定理

結果 E が起ったときに原因が A_i である確率は

$$P(A_i \mid E) = \frac{P(A_i)P(E \mid A_i)}{P(A_1)P(E \mid A_1) + P(A_2)P(E \mid A_2) + \cdots + P(A_n)P(E \mid A_n)}$$

Thomas Bayes
1702 –1761

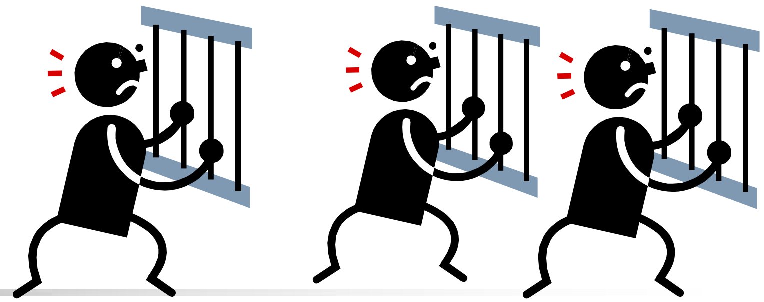
ベイズの定理:例題



- 同一の製品をつくっている３個の機械A,B,Cがある。
- A,B,Cは全製品のそれぞれ30%, 25%, 45%を生産する。
- A,B,Cの製品中の不良品の割合は,それぞれ1%, 1.2%, 2%であるとする。
- いま,全製品中から1個の製品をとりだしたとき,それが不良品であったという。
- この製品がA,B,Cのそれぞれの機械から生産された確率を求めよ.

Monty Hall問題

3囚人の問題



•3人の囚人A、B、Cがいる。三人とも処刑されることになっていたが、一人だけ恩赦されることになった。誰が恩赦になるのか決定されたが、まだ囚人たちには知らされていない。
•結果を知っている看守に、囚人Aが『BとCのうち、どちらかは必ず処刑されるのだから、処刑される一人の名前を教えてくれても、私に情報を与えることにはならないだろう。一人を教えてくれないか。』と頼んだ。
•看守は、その言い分に納得して『囚人Bは処刑されるよ』と教えてやった。
•囚人Aは、「はじめ自分の助かる確率は1/3だった。いまや助かるのは自分とCだけになったので、助かる確率は1/2になった」と喜んだ。

ベイズの定理

結果 E が起ったときに原因が A_i である確率は

$$P(A_i \mid E) = \frac{P(A_i)P(E \mid A_i)}{P(A_1)P(E \mid A_1) + P(A_2)P(E \mid A_2) + \cdots + P(A_n)P(E \mid A_n)}$$

Thomas Bayes
1702 –1761

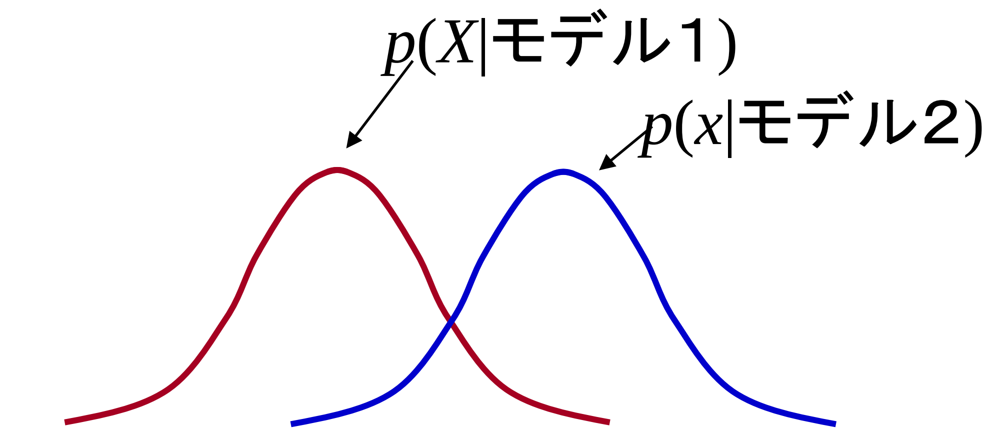
ベイズの定理

データXのモデルに基づく
生成率

モデルΘの事前確率

$$p(\Theta|X) = \frac{p(X|\Theta)p(\Theta)}{p(X)} = \frac{p(X|\Theta)p(\Theta)}{\sum_{\Theta'} p(X|\Theta')p(\Theta')}$$

モデルΘの事後確率
(データXを観測した後)



最尤推定（ML: Most likelihood）

$$\hat{\theta} = \arg\max_{\theta} \; p(y_o|\theta)$$

- 観測データの確率を最大にするようにパラメータを最適化する
- *y_o :* 観測データ
- *$p(y/\theta)$:* モデル
- θ: モデルのパラメータ

最尤推定（ML: Most likelihood）

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

- 例：コインが1つある。このコインを5回投げたところ、表、表、裏、表、裏と出た。このコインの表の出る確率をθとし、この値を推定せよ。

最尤推定（ML: Most likelihood）

- $P(D|\theta)$（尤度）
  - モデルパラメータ θ のもとでのデータ D の出現確率
- 最尤法
  - $P(D|\theta)$ を最大化する θ を選ぶ
- 例
  - コインを5回投げて、表が3回出た後、裏が2回出た
  - p(表)$=a$, p(裏)$=1-a$ とすると、$P(D|\theta)=a^3(1-a)^2$
  - $a=3/5$の時、$P(D|\theta)$ は最大
  - 一般に表が出る頻度を f とすると $a=f$ で尤度は最大

MAP推定(Maximum *A-Posteriori* Estimation)

$$\hat{\theta} = \mathop{\arg\max}_{\theta} \; p(\theta | y_o)$$ 式（1）

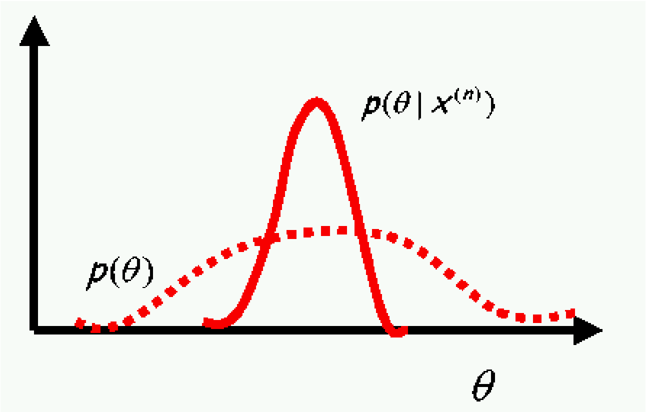
$$\hat{\theta} = \mathop{\arg\max}_{\theta} \; p(y_o | \theta)\, p(\theta)$$ 式（2）

- 観測データが与えられたときのパラメータの事後確率を最大にするパラメータを求める
- ベイズの定理から分母を無視すると、(2)が導かれる
- ***p(θ)***：事前確率

$$\Pr[\theta | X] = \frac{\Pr[X|\theta]\Pr[\theta]}{\int \Pr[X|\theta]\Pr[\theta] d\theta}$$

ベイズ推定

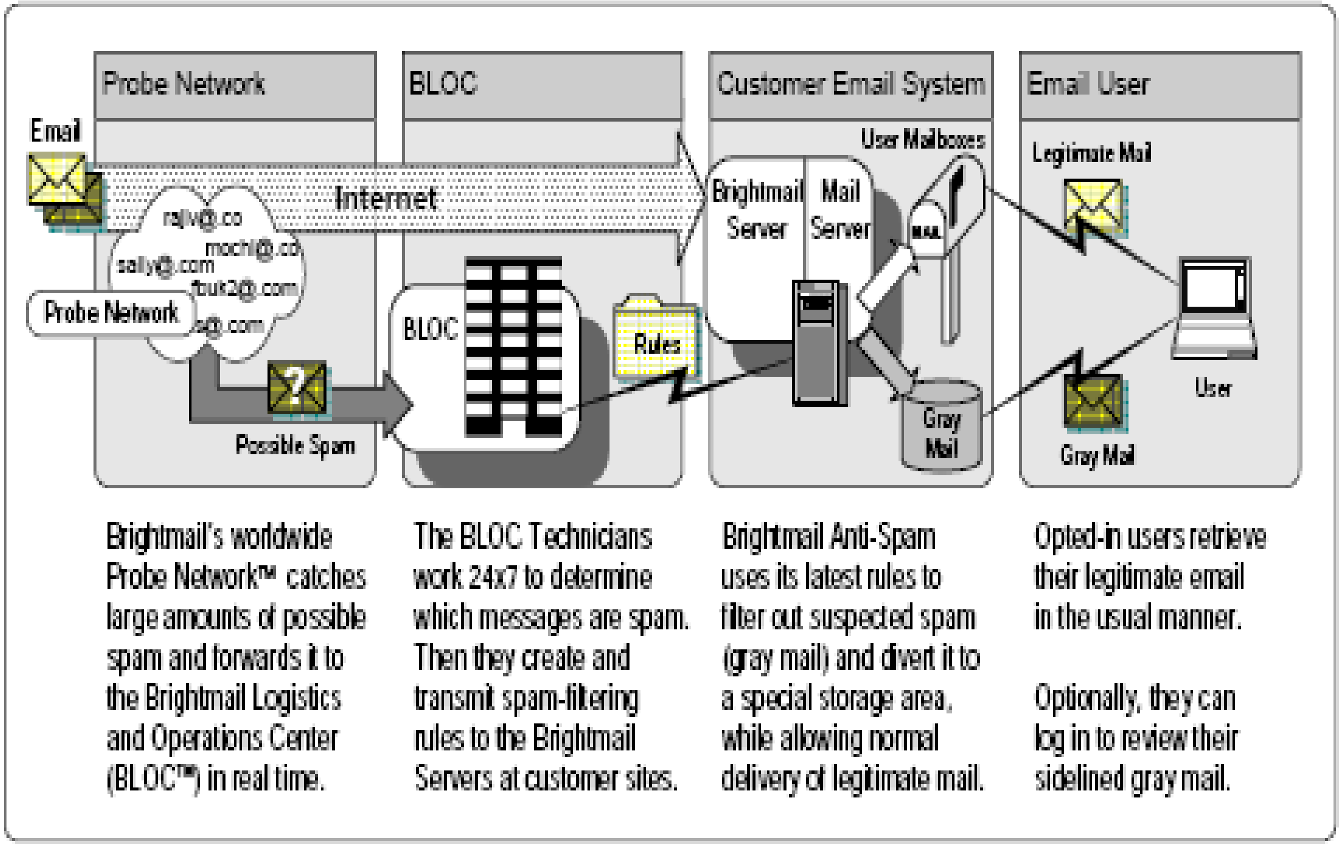
- パラメータθ を確率変数とみなして、パラメータの値の確信度を確率密度分布を用いて表現する。



- データ を観測する前にパラメータが取るであろう値の確率密度分布 を事前確率として表現し、データが観測された後にパラメータが取るであろう値の確率密度分布(事後確率密度分布) を推定する。

Bayesian Spam filter

- Mailwall processing flow [Brightmail02]



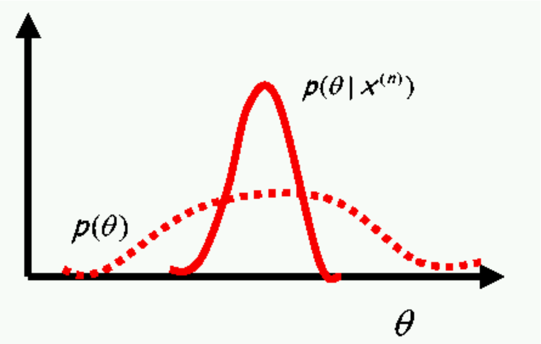
ベイズ推定

- パラメータθ を確率変数とみなして、パラメータの値の確信度を確率密度分布を用いて表現する。

- データ を観測する前にパラメータが取るであろう値の確率密度分布 を事前確率として表現し、データが観測された後にパラメータが取るであろう値の確率密度分布(事後確率密度分布) を推定する。

ベイズ推定

- データを観測する前にはパラメータがどんな値を取るかに関する情報が得られないので、パラメータの取る値の確率密度分布は、広がった分布となる。



- データが観測されると事後確率密度分布は、データと整合性の良いパラメータほど大きな値を持つような分布となる。
- つまり、事後確率分布は事前確率分布よりも狭い分布となる。

ベイズ推定

- N個の学習用データの集合 $X^{(n)} = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$ から事後確率密度分布 p(θ|X)が計算できるとする

- 学習用データと同じ分布から特徴ベクトル x が得られる確率密度分布は、

$$p(\mathbf{x} \mid X^{(n)}) = \int p(\mathbf{x}, \theta \mid X^{(n)}) d\theta$$

- 条件付き確率密度分布の定義から

$$p(\mathbf{x}, \theta \mid X^{(n)}) = p(\mathbf{x} \mid \theta, X^{(n)}) p(\theta \mid X^{(n)}).$$

ベイズ推定

$$p(\mathbf{x},\theta \mid X^{(n)}) = p(\mathbf{x} \mid \theta, X^{(n)})\, p(\theta \mid X^{(n)}).$$

- このとき、xはパラメータθのみに依存し、データX に依存しない。つまり、 $p(\mathbf{x} \mid \theta, X^{(n)}) \Rightarrow p(\mathbf{x} \mid \theta)$ とする。

- よって、最初の式は次のようになる。

$$p(\mathbf{x} \mid X^{(n)}) = \int p(\mathbf{x} \mid \theta)\, p(\theta \mid X^{(n)})\, d\theta$$

ベイズ推定

$$p(\mathbf{x} \mid X^{(n)}) = \int p(\mathbf{x} \mid \theta) p(\theta \mid X^{(n)}) d\theta$$

- つまり、ベイズ推定では、パラメータθ の特定の値を決める代わりに、すべての可能な値を考え、p(θ|X)を重みとした重み付き平均によりxの確率密度分布を推定する。

ベイズ推定とMAP推定

- ベイズ推定：尤度とモデル（パラメータ）の事前確率から、ベイズの定理により、事後確率を推定

$$P(\theta \mid D) = \frac{P(D \mid \theta) P(\theta)}{P(D)}$$

ただし、$P(D) = \int_{\theta'} P(D \mid \theta') P(\theta')$　（θが連続値の時）

- 最大事後確率（MAP）推定
  - $P(D|\theta)P(\theta)$ を最大化する θ を計算
  - $P(\theta)$ が一様分布なら最尤推定と同じ

不正サイコロのベイズ推定

- 公正サイコロと不正サイコロ
  - 公正: $P(i|\text{公正})=1/6$
  - 不正: $P(6|\text{不正})=1/2$, $P(i|\text{不正})=1/10$ for $i\neq6$
  - $P(\text{公正})=0.99$, $P(\text{不正})=0.01$

- ６が3回続けて出た場合の事後確率

$$P(\text{不正}\,|\,666)=\frac{P(666\,|\,\text{不正})P(\text{不正})}{P(666)}$$

$$=\frac{(0.5)^3(0.01)}{(0.5)^3(0.01)+(\frac{1}{6})^3(0.99)}=0.21$$