北数教　第 105 回数学教育実践研究会　平成 30 年 6 月 2 日（土）
北海道大学情報教育館

ベイズの定理からＡＩの数理

ベイズ更新とロジステック曲線について

松本睦郎（札幌啓成高等学校　講師）

Episode1 ロジステック曲線

　菌やウイルスの増殖数や、人口増加等を表現する曲線の一つにロジステック曲線があります。

例１　シャーレの中で培養された大腸菌の数について考察する。

シャーレ内に栄養が十分に存在するとき、菌は栄養を吸収しながら、一定時間ごとに細胞分裂をして増殖する。菌の数$u = u(t)$は時間の連続関数とみなすとき、増殖率は菌全体の数に比例する。

$$\frac{du}{dt} = au$$

（$a > 0$は比例定数）

微分方程式を解くと

$$\mathrm{u(t)} = u_0\, e^{at}$$

菌数は、指数関数的に増殖し、無限大に発散する。実際にはシャーレ内の栄養素が不足し環境悪化により、増殖率は低下する。分裂速度が u に比例して低下するものとすると、ロジステック微分方程式

$$\frac{du(t)}{dt} = (a - b\,u(t))u(t)$$

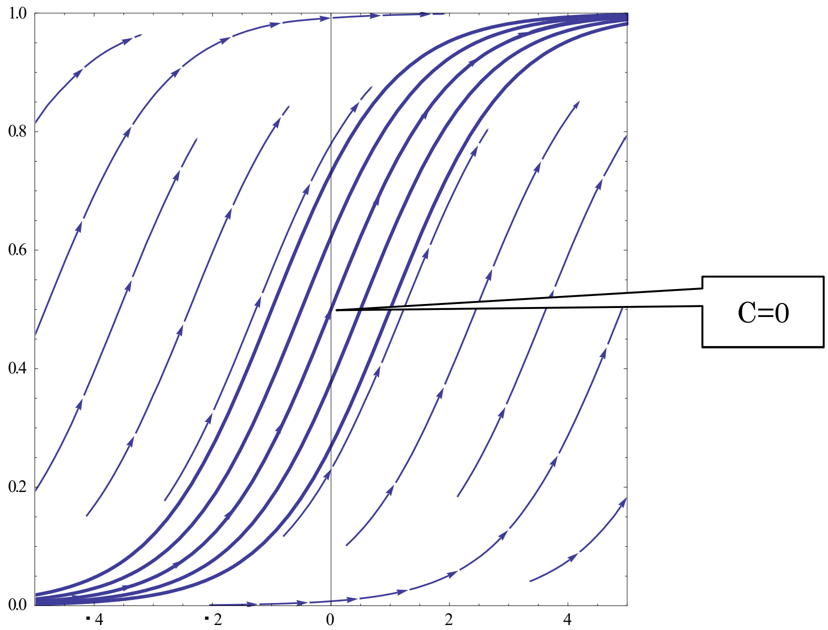
が得られる。$a = 1, b = 1$として、このベクトル場と解曲線を表示する。(↓)

簡単な変数分離型方程式を解くと

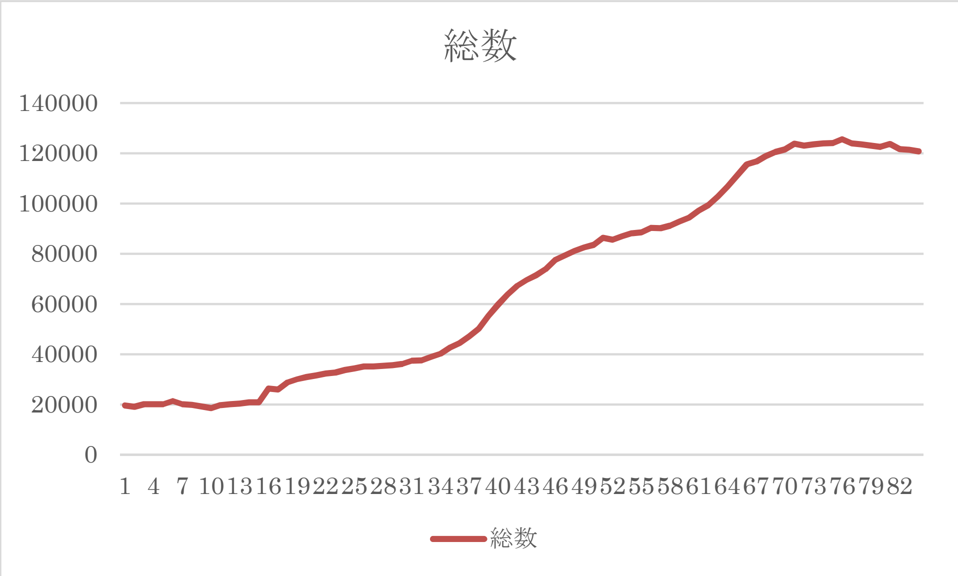
$$u(t) = \frac{e^{t+C}}{1 + e^{t+C}}$$

C は積分定数

C	$-\infty$	0	$+\infty$
$u(0)$	0	$\frac{1}{2}$	1



江別市の 1930 年から 2013 年までの人口推移は、ほぼロジステック曲線に従っている。



Episode 2　ロジステック回帰分析

不健康集団と健康集団を 0（健康）、1（不健康）に数値変換して、1 日の喫煙本数と一カ月間の飲酒日数を調査した。このデータについて、健康か不健康の判別モデルを作成するときロジステック回帰分析を使う。1 日の喫煙本数が 25 本、1 カ月の飲酒日数が 15 日である A さんの不健康有無を判別する。

データ

No	不健康有無	喫煙本数／日	飲酒日数／月
1	1	30	21
2	1	22	10
3	1	26	25
4	1	14	20
5	0	6	10
6	0	2	15
7	0	6	5
8	0	0	5
9	0	0	15
A さん	?	25	15

多変数ロジステック回帰方程式

$$z = \frac{1}{1 + e^{-(ax+by+c)}}$$

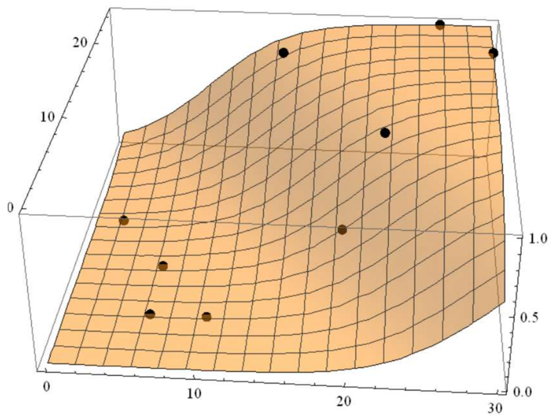
不健康有無を目的変数として z 軸、飲酒量（y 軸）と喫煙量（x 軸）を説明変数とする。

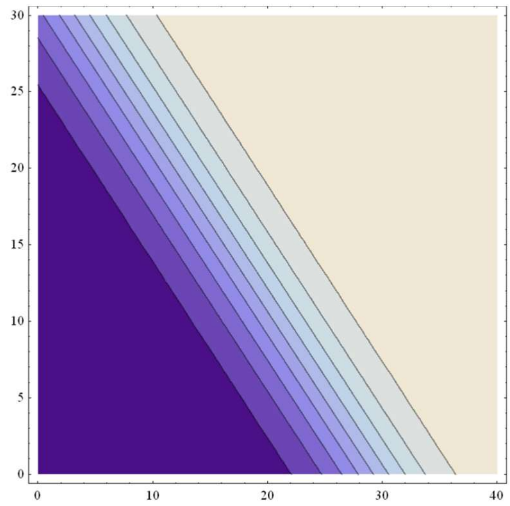
a, bを回帰係数、cを定数項と呼ぶ。

Mathematica10.0 を利用して回帰係数と、定数項を求める。

$$\mathrm{Plot3D}[\frac{1}{1+\mathrm{Exp}[8.99615-0.307914x-0.266886y]},\{x,0,30\},\{y,0,25\},\mathrm{Epilog}$$

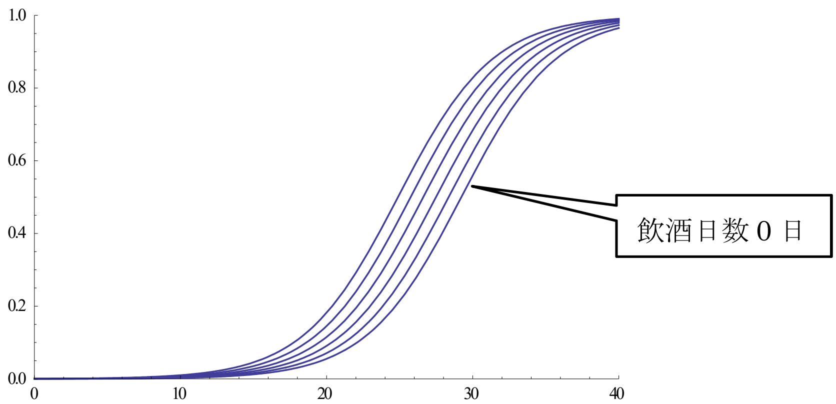
$$\rightarrow \{\mathrm{PointSize}[\mathrm{Medium}],\mathrm{Table}[\mathrm{Point}[\mathrm{data}[[i]]],\{i,1,9\}]\}]$$





hanbetu[x_,y_]:=1/(1+Exp[8.99615-0.307914 x-0.266886 y])

Plot[Table[hanbetu[x,i],{i,0,5}],{x,0,40},PlotRange→{{0,40},{0,1}}]



hanbetu[20,15]＝0.762316

1 日の喫煙本数が 20 本かつ 1 ケ月飲酒日が 15 日の人が、狭心症になる確率は、0.76 となる。
過去のデータから、未来の確率を求めることができる。

Episode 3　ベイズ定理

第 104 回数実研で札幌手稲高校　西村昴介先生の発表した「ベイズ統計学を用いた数学Aでの条件付き確率の導入の工夫」は、とても興味深いレポートでした。ベイズの定理を活用した例題について考えてみました。

【例題１】感染症の検査問題

インフルエンザのような感染症の疑いのある場合は、検査をおこなう。陽性と陰性の検査結果には誤差が発生する。今、p を感染している人が陽性となる確率、$1-p$ は感染している人が陰性となる確率。
q は健康な人が陽性となる確率、$1-q$ は健康な人が陰性となる確率とする。

（１）　ある人が検査を受けて陽性のとき、感染している確率を求めよ。

（２）　ある人が検査を受けて陽性のとき、感染していない確率を求めよ。

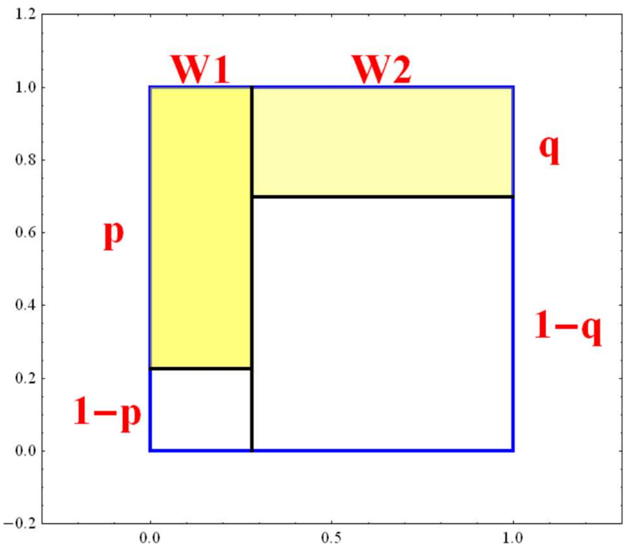
ただし、このインフルエンザ感染率を r として考えよ。

〔解答例〕（１）ベイズの定理を活用する。

W1：感染している事象、W2：健康である事象、B：陽性である事象、C：陰性である事象
とすると、条件付き確率より

$$P(W1)=r,P(W2)=1-r$$

$$P_{W1}(B)=p,P_{W1}(C)=1-p,P_{w2}(B)=q,P_{w2}(C)=1-q$$



$$P(B)=P(W1\cap B)+P(W2\cap B)\cdots①$$

$$P_{w1}(B)=\frac{P(w1\cap B)}{P(W1)}\quad \text{より}\quad P(W1\cap B)=P_{w1}(B)\times P(W1)=p\times r\cdots②$$

$$P_{w2}(B)=\frac{P(w2\cap B)}{P(W2)}\quad \text{より}\quad P(W2\cap B)=P_{w2}(B)\times P(W2)=q\times(1-r)\cdots③$$

②③を①へ代入すると、陽性の確率を求めることができる。

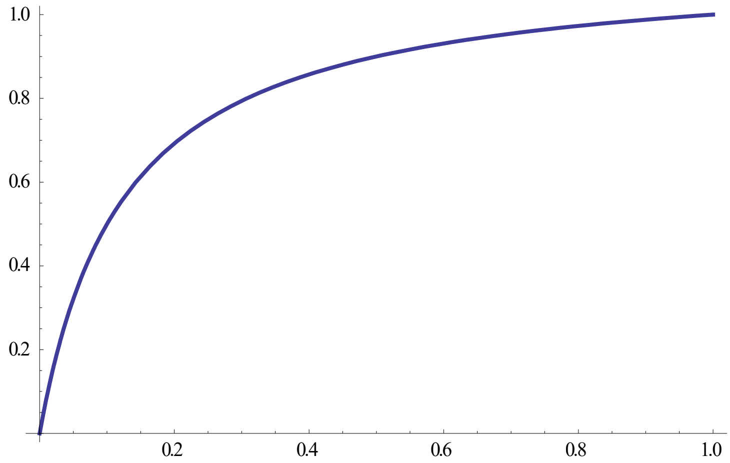
$$P(B)=(p-q)r+q\cdots④$$

陽性のとき、感染している確率は、

$$P_B(W1)=\frac{P(W1\cap B)}{P(B)}=\frac{p\times r}{(p-q)r+q}$$

$$p = 0.8, q = 0.1$$

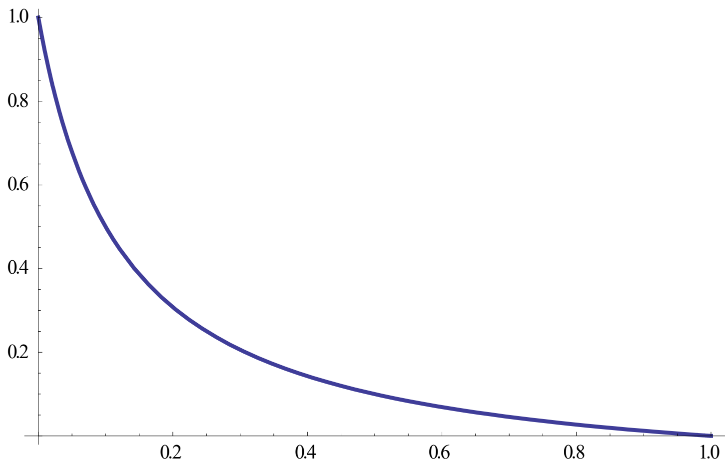
として、横軸に感染率（r）、縦軸に陽性にとき感染している（真陽性）の確率をとってグラフ化してみる。



感染確率や罹患確率が高いほど、真陽性の確率が高くなることがわかる。罹患確率が低いと、真陽性の確率も低い。

（２）陽性のとき感染していない確率は、

$$P_B(W1) = \frac{P(W2 \cap B)}{P(B)} = \frac{q \times (1 - r)}{(p - q)r + q}$$



感染確率や罹患確率が高いほど、偽陽性の確率が低くなることがわかる。罹患確率が低いと、偽陽性の確率も高くなる。

Episode 3　ベイズ更新とロジステック曲線

$P(W1) = r$を事前確率と呼ぶ。初めに

$$\mathrm{P}(W1) = 0.004 \cdots \text{（ア）}, P(w2) = 1 - 0.004 = 0.996$$

$$P_{W1}(B) = p = 0.8, P_{W1}(C) = 1 - p = 1 - 0.8 = 0.2 \,, P_{w2}(B) = q = 0.1, P_{w2}(C) = 1 - q = 0.9$$

と設定することにする。

陽性のとき罹患する確率は、

$$P_B(W1) = \frac{P(W1 \cap B)}{P(B)} = \frac{p \times r}{(p - q)r + q} = \frac{0.8 \times 0.004}{(0.8 - 0.1) \times 0.004 + 0.1} = 0.031128404 \cdots \text{（イ）}$$

真陽性の確率は、0.031（3.1%）となり、事後確率と呼ぶ。
ベイズ更新とは、ベイズの定理によって算出した事後確率と、次回の試行の事前確率としてもう一度ベイズの定理を活用することである。

$$\mathrm{P(W1)} = 0.031, \mathrm{P(W2)} = 1 - 0.031 = 0.969$$

を事前確率として、陽性のときの罹患している真陽性の確率は、

$$P_B(W1) = \frac{P(W1 \cap B)}{P(B)} = \frac{p \times r}{(p-q)r + q} = \frac{0.8 \times 0.031}{(0.8 - 0.1) \times 0.031 + 0.1} = 0.203779786 \cdots \text{（ウ）}$$

真陽性確率の推移は

$$0.004 \to 0.031128404 \to 0.203779786$$

（ｉ）陽性の場合の対数オッズ比

$$\log \frac{p}{q} = log 8 = 2.07944 \cdots ①$$

（ⅱ）陰性の場合のオッズ比

$$\log \frac{1-p}{1-q} = log \frac{2}{9} = -1.50408 \cdots ②$$

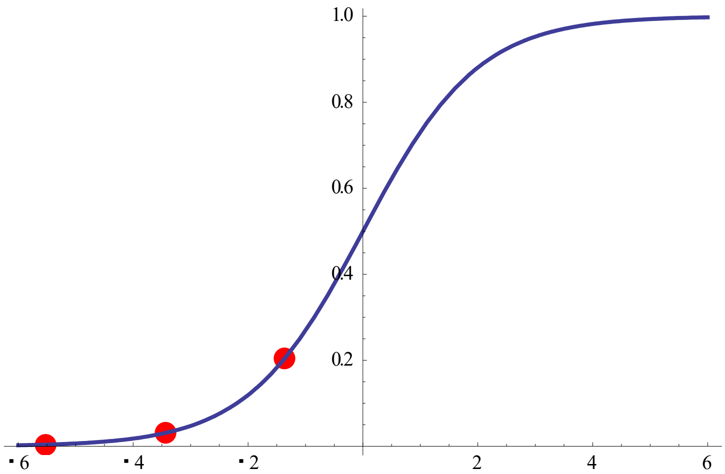
（ⅲ）スタートスコア

$$\log \frac{W1}{W2} = log \frac{0.004}{0.996} = -5.5174 \cdots ③$$

（ア）（イ）（ウ）①②③から何かが見える。
座標（③,（ア)),（③+①,（イ)),（③+２×①,（ウ)）を表示すると、

(−5.5174,0.004),(−3.43796,0.031),(−1.35852,0.203779786)

この３点を座標平面上に表示する。



これらの３点は、ロジステック曲線上にすべて存在する。ベイズ更新は、すべてこの曲線上の確率となる。毎年病気の検査で陽性判定が繰り返されとすると、真陽性の確率がこの曲線

$$y = \frac{1}{1 + e^{-x}}$$

に従って、上昇していくことになる。逆に陰性の場合 x 軸を②の割合で真陽性の確率が減少していく。
AI で利用されている数学のひとつである。

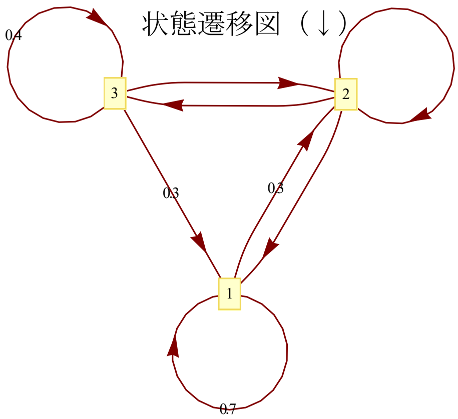
Episode4 マルコフ連鎖

映画「ジュラシック・パーク」の中で、非線形微分方程式を研究している数学者マルカム博士が登場する。「自然の中に内包する予測不可能性（カオス）によって失敗する。」セリフがある。マルカムはマルコフ連鎖を想像させた。マルコフ連鎖とは、「未来の確率が現在の状態のみで決定される。過去のいかなる情報も、未来を予測する際には無関係である。」

【例題 2】天気の遷移確率

昨日以前の天気は翌日の天気に影響しない。今日「晴れ」のとき明日「晴れ」の確率は、0.7、「曇り」の確率は 0.3、「雨」の確率は 0。今日「曇」のとき明日「晴れ」の確率は、0.4、「曇り」の確率は 0.4、「雨」の確率は 0.2。今日「雨」のとき明日「晴れ」の確率は、0.3、「曇り」の確率は 0.3、「雨」の確率は 0.4。状態空間は、S={晴，曇，雨}である。晴→1、曇→2、雨→3 で表示すると、S={1,2,3}となる。

状態遷移図（↓）



X_t ：t 日目の天気を表す確率変数とすると、条件付き確率より t 日目に晴れたとき、$t+1$ 日目も晴れる確率は、$P_{xt=1}(X_{t+1} = 1) = 0.7$

推移確率行列とは(i,j)成分に i から j に遷移する確率を行列で表示する。

$$P = \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

性質ｉ：推移確率行列の各要素は、0 以上 1 以下

性質ⅱ：各行の和は、1 となる。

$P_{xt=i}(X_{t+n} = i)$を(i,j)成分の行列つまり、n日経過の推移確率行列を$P^{(n)}$とする。チャップマン・コルモゴロフ方程式：$P^{(n)} = P^n$ により、推移確率行列の n 乗とn日経過の推移確率行列が一致する。

$$P^2 = \begin{pmatrix} 0.48999999999999994 & 0.09 & 0 \\ 0.16000000000000003 & 0.16000000000000003 & 0.04000000000000001 \\ 0.09 & 0.09 & 0.16000000000000003 \end{pmatrix}$$

$$P^3 = \begin{pmatrix} 0.3429999999999999 & 0.027 & 0 \\ 0.06400000000000002 & 0.06400000000000002 & 0.008000000000000002 \\ 0.027 & 0.027 & 0.06400000000000002 \end{pmatrix}$$

AI 人工知能というフレーズが最近良く耳にする。決して AI は絶対的万能なものではなく、統計学（数理統計、ベイズ統計）や確率論（確率過程）等の従来の学問を基本にしたものであり、確率的なものであることがわかる。　（おわり）