ベイズ推定（基本編）

1 ベイズの定理

確率に関するベイズの定理を復習する．

定義（条件付き確率）

事象A_iとB_jを考え，A_iとB_jが同時に起こる確率を$p(A_i, B_j)$とする．そして，

$$p(B_j|A_i) = \frac{p(A_i, B_j)}{p(A_i)}$$

とする．$p(B_j|A_i)$を，A_iという条件のもとでB_jが起こる**条件付き確率**をという．

この定義より，

$$p(A_i, B_j) = p(B_j|A_i)p(A_i)$$

がいえ，同様に，

$$p(A_i, B_j) = p(A_i|B_j)p(B_j)$$

もいえる．したがって，次の定理が得られる．

$$p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{p(B_j)} \tag{1}$$

を得る．このとき，$P(A_i)$をA_iの**事前確率**，$P(A_i|B_j)$をA_iの**事後確率**という．また，$p(B_j|A_i)$を **尤度**（ゆうど） という．

さて，

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i, B_j) = 1$$

とする．このとき，

$$p(B_j) = \sum_{i=1}^{n} p(A_i, B_j) = \sum_{i=1}^{n} p(B_j|A_i)p(A_i) \tag{2}$$

であり，(1)と(2)より，以下の定理が得られる．

確率に関するベイズの定理

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(A_i, B_j) = 1 \text{ ならば} \quad p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{\sum_{i=1}^{n} p(B_j|A_i)p(A_i)} \tag{3}$$

[例題 1] ある工場の製品は，機械M_1では単位時間当たり 15 個，機械M_2では単位時間当たり 20 個の速さで製造されている．しかし，M_1では3%の不良品が，M_2では5%の不良品が含まれる．いま，一定の時間にM_1とM_2により製造された製品の中から任意に取り出した 1 個が不良品であった．この不良品がM_1の製造品である確率を求めよ．

（解答）単位時間において，事象A_1をM_1の製造品，事象A_2をM_2の製造品とする．さらに，事象Bを不良品とする．このとき，

$$P(A_1)=\frac{15}{35},\qquad P(A_2)=\frac{20}{35},\qquad P(B|A_1)=\frac{3}{100},\qquad P(B|A_2)=\frac{5}{100}$$

である．これより，

$$p(B|A_1)p(A_1)=\frac{3}{100}\times\frac{15}{35}=\frac{9}{700},\qquad p(B|A_1)p(A_1)+p(B|A_2)p(A_2)=\frac{3}{100}\times\frac{15}{35}+\frac{5}{100}\times\frac{20}{35}=\frac{29}{700}$$

よって，求める確率である$P(A_1|B)$は，ベイズの定理より

$$P(A_1|B)=\frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1)+p(B|A_2)p(A_2)}=\frac{9}{29}$$

となる．□

2 ベイズ推定

事象B_jとC_kは互いに独立，すなわち，$p(B_j,C_k)=p(B_j)p(C_k)$の場合には，

$$p(A_i,B_j,C_k)=p(A_i|B_j,C_k)p(B_j,C_k)$$
$$p(A_i,B_j,C_k)=p(B_j,C_k|A_i)p(A_i)$$

が成り立つので，

$$p(A_i|B_j,C_k)=\frac{p(B_j,C_k|A_i)p(A_i)}{p(B_j,C_k)}=\frac{p(C_k|A_i)}{p(C_k)}\cdot\frac{p(B_j|A_i)p(A_i)}{p(B_j)}=\frac{p(C_k|A_i)p(A_i|B_j)}{p(C_k)}\tag{4}$$

が得られる．ここで，B_jとC_kは互いに独立より，

$$p(C_k|A_i)=p(C_k|A_i,B_j)=\frac{p(A_i,B_j,C_k)}{p(A_j,B_j)}=\frac{p(A_i,C_k)p(B_j)}{p(A_i|B_j)p(B_j)}=\frac{p(A_i,C_k)}{p(A_i|B_j)}$$

であることに注意すると，

$$p(C_k)=\sum_{i=1}^{n}p(A_i,C_k)=\sum_{i=1}^{n}p(C_k|A_i)p(A_i|B_j)$$

となる．よって，(4)より次の定理を得る．

定理（**ベイズ推定**，または**ベイズ更新**）事象B_jとC_kは互いに独立ならば，

$$p(A_i|B_j,C_k)=\frac{p(C_k|A_i)p(A_i|B_j)}{\sum_{i=1}^{n}p(C_k|A_i)p(A_i|B_j)}\tag{5}$$

[例題 4.1] 迷惑メールフィルターは，迷惑メールA_1か一般メールA_2を振り分けるものである．メールAには「無料」,「未納料金」,「サイト利用料金」という言葉が含まれている．迷惑メールと一般メールに，それらが含まれる確率は，これまでのデータベースから，以下と表のとおりである．また，データベースから現時点での迷惑メールの割合は 6 割であることもわかっている．

	迷惑メールA_1	一般メールA_2
無料	0.12	0.02
未納料金	0.15	0.01
サイト利用料金	0.25	0.02

無料という言葉が含まれている事象をB_1，そうでない事象をB_2とし，未納料金という言葉が含まれている事象をC_1，そうでない事象をC_2とし，サイト利用料金という言葉が含まれている事象をD_1，そうでない事象をD_2とする．このとき，$p(A_1|B_1, C_1, D_1)$をベイズ更新から推定せよ．

（解答）表より，$p(B_1|A_1) = 0.12$である．また，$p(A_1) = 0.6$である．(4.3)より，

$$p(A_1|B_1) = \frac{p(B_1|A_1)p(A_1)}{p(B_1|A_1)p(A_1) + p(B_1|A_2)p(A_2)} = \frac{0.12 \times 0.6}{0.12 \times 0.6 + 0.02 \times 0.4} = 0.9000$$

である．つぎに，$p(C_1|A_1) = 0.15$であることから，(4.5)より

$$p(A_1|B_1, C_1) = \frac{p(C_1|A_1)p(A_1|B_1)}{p(C_1|A_1)p(A_1|B_1) + p(C_1|A_2)p(A_2|B_1)} = \frac{0.15 \times 0.9}{0.15 \times 0.9 + 0.01 \times 0.1} = 0.9926$$

となる．最後に，$p(D_1|A_1) = 0.25$であることから，(4.5)より

$$p(A_1|B_1, C_1, D_1) = \frac{p(D_1|A_1)p(A_1|B_1, C_1)}{p(D_1|A_1)p(A_1|B_1, C_1) + p(D_1|A_2)p(A_2|B_1, C_1)} = \frac{0.25 \times 0.9926}{0.25 \times 0.9926 + 0.02 \times 0.0074} = 0.9993$$

以上より，メールAは迷惑メールである可能性が極めて高い．□

3 頻度確率とベイズ確率の考え方の違い

現代の統計学では，大きく頻度確率とベイズ確率と呼ばれる 2 種類の考え方で扱われる．以下，簡単にそれらを説明する．

統計を頻度確率で扱おうとする考え方は，できるだけ偏りのない観測を多数得ることで，公平・公正な判断を導くことに重きを置くというものである．したがって，多少手間をかけても良いから，より精確なデータ解析を必要とする用途に適している．しかし，欲しい情報を欲しいだけ観測できるかどうかは問題である．

統計をベイズ確率で扱おうとする考え方は，多少データが偏っていてもよいから，少ない観測で効率的に推定が行えることを優先するものである．そのためなら，外部の情報を事前知識として持ち込むことを厭わず，使えるものは使いたいとする考え方である．しかし，事前知識が不正確だと推定は怪しくなるというリスクは必然である．それでもそのリスクを負うことを承知してでも，推定値を効率的に得えようとする考え方である．具体的には，観測自体ができない状態を扱いたい場合，ある種の仮説を立てて，演繹的な推論を行うのである．

4 確率分布に関するベイズの定理とベイズ推定

θとxを確率変数とする．$f(\theta,x)$をθとxに関する確率密度関数とする．さらに，$f(x|\theta)$をθを固定したときのxに関する確率密度関数とする．確率に関する

$$p(A_i,B_j)=p(B_j|A_i)p(A_i),\qquad p(A_i,B_j)=p(A_i|B_j)p(B_j)$$

から，$f(\theta,x)$と$f(x|\theta)$との間に

$$f(\theta,x)=f(x|\theta)f(\theta),\qquad f(\theta,x)=f(\theta|x)f(x) \tag{6}$$

という関係式を定義する．これより，

$$f(\theta|x)=\frac{f(x|\theta)f(\theta)}{f(x)} \tag{7}$$

を得る．$f(\theta)$をθの**事前確率分布**，$f(\theta|x)$をθの**事後確率分布**という．また，$f(x|\theta)$をxの**尤度関数**という．

また，確率に関する式

$$p(B_j)=\sum_{i=1}^{n}p(B_j|A_i)p(A_i)$$

に対応させて

$$f(x)=\int_{-\infty}^{\infty}f(x|\theta)f(\theta)d\theta \tag{8}$$

を定義する．

(8)と(9)から，**確率分布に関するベイズの定理**

$$f(\theta|x)=\frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{\infty}f(x|\theta)f(\theta)d\theta} \tag{9}$$

が得られる．x_1,x_2が互いに独立であるとき，ベイズ更新

$$p(A_i|B_j,C_k)=\frac{p(C_k|A_i)p(A_i|B_j)}{\sum_{i=1}^{n}p(C_k|A_i)p(A_i|B_j)}$$

に対応して，

$$f(\theta|x_1,x_2)=\frac{f(x_2|\theta)f(\theta|x_1)}{\int_{-\infty}^{\infty}f(x_2|\theta)f(\theta)d\theta} \tag{10}$$

となる．さらに，

$$k_2=\frac{1}{\int_{-\infty}^{\infty}f(x_2|\theta)f(\theta)d\theta}$$

と置くと，(10)は

$$f(\theta|x_1,x_2)=k_2f(x_2|\theta)f(\theta|x_1) \tag{11}$$

となる．これを一般化すると以下となる．

$$f(\theta|x^{(n)})=k_nf(x_n|\theta)f(\theta|x^{(n-1)}),\qquad \frac{1}{k_n}=f(x^{(n)})=\int_{-\infty}^{\infty}f(x_n|\theta)f(\theta)d\theta \tag{12}$$

以上をまとめると，以下となる．

確率分布に関するベイズ推定は

$$(1) \quad f(\theta|x_1) = k_1 f(x_1|\theta) f(\theta)$$

$$(2) \quad f(\theta|x_1, x_2) = k_2 f(x_2|\theta) f(\theta|x_1)$$

$$(3) \quad f(\theta|x_1, x_2, x_3) = k_3 f(x_3|\theta) f(\theta|x_1, x_2)$$

$$\cdots\cdots\cdots\cdots$$

$$(4) \quad f\left(\theta|x^{(n)}\right) = k_n f(x_n|\theta) f\left(\theta|x^{(n-1)}\right) \quad \text{ただし，} x^{(n)} = x_1, x_2, \cdots, x_n$$

となる．

[例題 3] 表の出る確率がθであるコインがある．このコインを 3 回投げたとき，表，表，裏の順に出た．このときにθ関する確率分布$f\left(\theta|x^{(3)}\right)$を推定せよ．ただし，事前分布は$f(\theta) = 1$とせよ．

（解答） 1 回目は表であるので$f(x_1|\theta) = \theta$であり，$f(\theta) = 1$より

$$f(\theta|x_1) = k_1 f(x_1|\theta) f(\theta) = k_1 \theta$$

である．$\int_0^1 f(\theta|x_1) d\theta = 1$と

$$\int_0^1 k_1 \theta d\theta = \frac{k_1}{2} [\theta^2]_0^1 = \frac{k_1}{2}$$

より，$k_1 = 2$を得る．よって，

$$f(\theta|x_1) = 2\theta$$

となる．次に，２回目も表より$f(x_2|\theta) = \theta$であり，$f(\theta|x_1) = 2\theta$より

$$f\left(\theta|x^{(2)}\right) = k_2 f(x_2|\theta) f(\theta|x_1) = 2k_2 \theta^2$$

であり，$\int_0^1 f(\theta|x_2) d\theta = 1$と

$$\int_0^1 2k_2 \theta^2 d\theta = \frac{2k_2}{3} [\theta^3]_0^1 = \frac{2k_2}{3}$$

より，$k_2 = 3/2$を得る．よって，

$$f\left(\theta|x^{(2)}\right) = \frac{3}{2} \cdot 2\theta \cdot \theta = 3\theta^2$$

となる．最後に，3 回目は裏より$f(x_3|\theta) = 1 - \theta$であり，$f\left(\theta|x^{(2)}\right) = 3\theta^2$より

$$f\left(\theta|x^{(3)}\right) = k_3 f(x_3|\theta) f\left(\theta|x^{(2)}\right) = 3k_3 \theta^2 (1 - \theta)$$

であり，$\int_0^1 f(\theta|x_2) d\theta = 1$と

$$\int_0^1 3k_3 \theta^2 (1 - \theta) d\theta = 3k_3 \left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^1 = \frac{k_3}{4}$$

より，$k_3 = 4$を得る．よって，θ関する確率分布$f\left(\theta|x^{(3)}\right)$は，

$$f\left(\theta|x^{(3)}\right) = 4 \cdot 3\theta^2 \cdot (1 - \theta) = 12\theta^2 (1 - \theta)$$

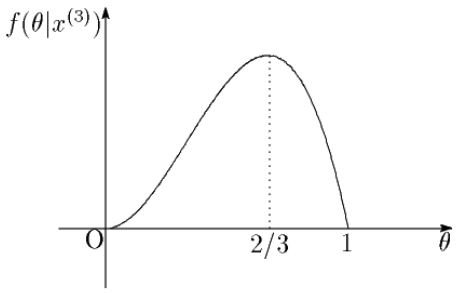
となる．

（解説）θ関する確率分布$f\left(\theta|x^{(3)}\right)=12\theta^2(1-\theta)$を考える．

$$f'=12\theta(2-3\theta)$$

$0<\theta<1$より，$\theta=2/3$で$f\left(\theta|x^{(3)}\right)$は極大値をとる．グラフは右図である．

そして，３個の統計から得られたこのコインの表の出る確率は，事前分布を一様分布として仮定した場合，$\theta=2/3$が確率的に最も高いということを意味するのである．



練習問題

[1] ある国民の0.02%がある病気に罹患している．この病気のある検査方法では，実際に病気に罹患している人が陽性と判定される確率は90%で，病気に罹患していない人が陰性と判定される確率は80%である．ある人がこの病気の検査を受けて陽性であった．この人が病気に罹患している確率（%）を，小数点以下第３位まで求めよ．

[2] ある記憶デバイスは，A社，B社，C社で製造されている．このデバイスの市場のシェアは，それぞれ50%, 30%, 20%である．また，このデバイスが不良品として返品される確率は，それぞれ2%, 5%, 9%である．ある人がこのデバイスを購入したら不良品であった．このデバイスがA社の製品である確率（%）を，小数点以下第３位まで求めよ．

[3] 表の出る確率がθであるコインがある．このコインを３回投げたとき，裏，表，裏の順に出た．このときにθ関する確率分布$f\left(\theta|x^{(3)}\right)$を推定せよ．ただし，事前分布は$f(\theta)=1$とせよ．また，$f\left(\theta|x^{(3)}\right)$の極大値をとる$\theta$を求めよ．

練習問題の答え［1］0.023%　［2］0.233%　［3］$f\left(\theta|x^{(3)}\right)=12\theta(1-\theta)^2$，$\theta=\frac{1}{3}$