

10

● 目次

★問題の頁数のあとのマス目は，自分の理解の度合いを記入しておくのにご利用ください.

まえがき	i
------	---

Chapter 1. 概要	1
---------------	---

問題 01 [標準] ベイズの定理 (Bayes' Theorem)	2 □□□
--	-------

問題 02 [やや難] 正規分布 $N(\mu, 1)$ の 2 乗の分布	4 □□□
---	-------

Tea Time ● 実験の計画が重要である：円周率の推定の場合	6
--	---

Chapter 1

概要

□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□

2 Chapter.1 概要

問題 01 ベイズの定理 (Bayes' Theorem)

標準

有望と思えるガン発見のある検査法が開発された。大病院のガン患者の 97% がこの検査に陽性反応を示し、ガンでない患者の 5% が同じ陽性反応を示したとしよう。病院の患者の 2% がガンにかかっているとするとき、無作為に選んだある患者がこの検査に陽性反応を示したとして、以下の問いに答えよ。

- (1) 抽出された患者が本当にガンにかかっている確率を求めよ。
- (2) (1) で得られた結果について吟味せよ。

この問題はベイズの定理 (Bayes' Theorem) に対する理解を求めている。

解説 まず、ベイズの定理を述べよう。

Ω は標本空間で、 H_1, H_2, \dots, H_k は Ω の部分集合である。標本空間 Ω は H_1, H_2, \dots, H_k によって素分割されている。すなわち、次が成り立つ。

- (1) $\Omega = H_1 \cup H_2 \cup \dots \cup H_k$ であり、
- (2) H_1, H_2, \dots, H_k は互いに排反である。すなわち、 $H_i \cap H_j = \phi (i \neq j)$ である。

このとき、任意の事象 A と任意 $i = 1, \dots, k$ に対して、次が成り立つ。

$$P(H_i|A) = \frac{P(H_i) P(A|H_i)}{\sum_{j=1}^k P(H_j) P(A|H_j)} \quad (1.1)$$

ベイズの定理は次のように証明される。まず、事象 A は Ω の部分集合なので、集合の吸収則により、 $A = A \cap \Omega$ となる。条件により、 $\Omega = H_1 \cup H_2 \cup \dots \cup H_k$ なので、集合の分配則により、

$$A = (A \cap H_1) \cup (A \cap H_2) \cup \dots \cup (A \cap H_k)$$

が成り立つ。確率の加法性により、 A の確率は

$$P(A) = \sum_{j=1}^k P(A \cap H_j) = \sum_{j=1}^k P(H_j) P(A|H_j) \quad (1.2)$$

と分解できる。2 番目の等号は乗法定理 $P(A \cap B) = P(B) P(A|B)$ による。(1.2) は全確率の定理 (law of total probability) として知られる。条件付確率の定義に

より,

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

と書ける。上の式の分母に全確率の定理を適用すれば、ベイズの定理 (1.1) が得られる。

解答

- (1) 患者の集団を Ω とし、ガン患者の集団を H_1 、ガンでない患者の集団を H_2 とすると、 Ω は H_1 と H_2 に素分割される。無作為に選んだ患者が陽性を示すという事象を A とすると、求めるべき確率は $P(H_1|A)$ と表現できる。以上の記号の下で、与えられた条件は以下のように整理できよう。

事前確率	尤度
$P(H_1) = 0.02$	$P(A H_1) = 0.97$
$P(H_2) = 0.98$	$P(A H_2) = 0.05$

したがって、全確率は

$$P(A) = P(H_1) P(A|H_1) + P(H_2) P(A|H_2) = 0.02 \times 0.97 + 0.98 \times 0.05 = 0.0684$$

となる。ベイズの定理より

$$P(H_1|A) = \frac{P(A|H_1) P(H_1)}{P(A)} = \frac{0.02 \times 0.97}{0.0684} \approx 0.284 = 28.4\%$$

- (2) ベイズの定理における $P(H_i)$ は事前確率 (prior probability) と呼ばれ、これまでの経験などに基づいて、病気などにおける確率に関する先験的 (prior) 知識を表す。新たなデータ A が得られたとき、病気になる事後確率 (posterior probability) $P(H_i|A)$ を計算しているのがベイズの定理である。より広い文脈において、 H_i を原因、 A を結果とすると、ベイズの定理は、結果が得られたときの原因の確率 $P(H_i|A)$ を更新するための理論として非常に重要である。上の例において、 $P(H_1|A) = 28.4\%$ がそれほど高くないことに意外と感じるかもしれない。これがむしろ一般的な現象である。もし、 $\Omega = H_1 \cup H_2$ で、 $P(H_1)$ が非常に小さければ、 $P(H_2) = 1 - P(H_1) \approx 1$ となる。 $P(A) \approx P(A|H_2)$ となることに注意すれば、 $P(H_1|A) \approx 0 / P(A|H_2) = 0$ となることが理解できよう。

4 Chapter.1 概要

問題	02	正規分布 $N(\mu, 1)$ の 2 乗の分布	やや難
<p>湖に野生してる蓮の花がそよ風に揺れ、時折不規則的に水面に触れる。水面から 25 センチ突き出ている蓮の花をじっと見ていたら、53 センチ前後離れた水面に触れることに段々と確信した。しかし、離れる距離は目測であり、誤差が伴うことも承知する。この誤差は、平均が 0 センチ、標準偏差が 5 センチの正規分布に従うとしよう。</p> <p>(1) 水の深さの期待値は何センチか。</p> <p>(2) 水の深さが 40 センチ以上となる確率はいくらか。</p>			

確率変数の関数の分布や積率母関数の求め方を理解することが前提である。

解説 $X \sim N(\mu, 1)$ とする。 $\mu = 0$ のとき、すなわち、 X が標準正規分布 $N(0, 1)$ に従うとき、 X^2 は自由度 1 のカイ自乗分布に従う。ここで、 $-\infty < \mu < \infty$ として、 $Y = X^2$ の密度関数 $f(y)$ と積率母関数 $M(t)$ を求めてみよう。

$\Phi(\cdot)$ を $N(0, 1)$ の分布関数とする。 $X - \mu \sim N(0, 1)$ に注意すると、任意の $y > 0$ に対して、 Y の分布関数は

$$\begin{aligned} F(y) &= \Pr[Y \leq y] = \Pr[X^2 \leq y] \\ &= \Pr[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= \Pr[-\sqrt{y} - \mu \leq X - \mu \leq \sqrt{y} - \mu] \\ &= 1 - \Phi(-\sqrt{y} - \mu) + \Phi(\sqrt{y} - \mu) \end{aligned}$$

となり、 $Y = X^2$ の密度関数 $f(y) = F'(y)$ は次のように書ける。

$$f(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}(y+\mu^2)} \cosh(\mu\sqrt{y}) \quad (y > 0) \quad (1.3)$$

ただし、 $\cosh x = (e^x + e^{-x})/2$ は双曲線余弦関数である。一方、積率母関数は次のように計算できる。

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tY}] = \mathbb{E}[e^{tX^2}] \\ &= \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{1-2t}} e^{\frac{t\mu^2}{1-2t}} \quad (t < 1/2) \end{aligned} \quad (1.4)$$

$M(t)$ を t について微分することにより、 Y の 1 次と 2 の積率を求めることができる。

$$\mathbb{E}(Y) = M'(0) = \mu^2 + 1 \quad (1.5)$$

$$\mathbb{E}(Y^2) = M''(0) = \mu^4 + 6\mu^2 + 3 \quad (1.6)$$

解答

- (1) 蓮の花の長さは、水面に出ている部分 $a = 25(\text{cm})$ と、水の深さ $Y(\text{cm})$ の和からなる。条件により、蓮の花が水面に触れるときの距離を $W(\text{cm})$ とすると、

$$W = b + \text{誤差} = b + N(0, \sigma^2) = N(b, \sigma^2)$$

となる。ただし、 $b = 53(\text{cm})$, $\sigma = 5(\text{cm})$ である。三平方の定理

$$(Y + a)^2 = Y^2 + W^2$$

により、 $Y = W^2/(2a)$ となる。 $X = W/\sigma$, $\mu = b/\sigma$ とすると、 $X \sim N(\mu, 1)$,

$$Y = \frac{\sigma^2}{2a} X^2 \quad (1.7)$$

となる。(1.5) により、水の深さ Y の期待値は

$$\mathbb{E}(Y) = \frac{\sigma^2}{2a} \mathbb{E}(X^2) = \frac{\sigma^2}{2a} \left(\frac{b^2}{\sigma^2} + 1 \right) = \frac{b^2 + \sigma^2}{2a} \quad (1.8)$$

となる。具体的数値を代入すると、 $\mathbb{E}(Y) = 56.68(\text{cm})$ となる。測定誤差がなければ、 $\sigma = 0$ となり、 $\mathbb{E}(Y) = b^2/(2a)$ となる。これが「ハスの問題」¹ として知られている。

- (2) 次に、 $Y \geq c = 40(\text{cm})$ の確率を求める。 $\lambda = 2ac/\sigma^2$ とすると、(1.7), (1.3) より、

$$\begin{aligned} \mathbb{P}[Y \geq c] &= \mathbb{P}[X^2 \geq \lambda] \\ &= \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}(y+\mu^2)} \cosh(\mu\sqrt{y}) dy \\ &= \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}(y+\mu^2)} \frac{e^{\mu\sqrt{y}} + e^{-\mu\sqrt{y}}}{2} dy \\ &= 2 - \Phi(\sqrt{\lambda} + \mu) - \Phi(\sqrt{\lambda} - \mu) \\ &= 2 - \Phi\left(\frac{\sqrt{2ac}}{\sigma} + \frac{b}{\sigma}\right) - \Phi\left(\frac{\sqrt{2ac}}{\sigma} - \frac{b}{\sigma}\right) \\ &= 2 - \Phi\left(\frac{\sqrt{2ac} + b}{\sigma}\right) - \Phi\left(\frac{\sqrt{2ac} - b}{\sigma}\right) \end{aligned}$$

ここで、 $\Phi[(\sqrt{2ac} + b)/\sigma] \approx 1$ に注意すると、

$$\mathbb{P}[Y \geq c] \approx 1 - \Phi\left(\frac{\sqrt{2ac} - b}{\sigma}\right) \quad (1.9)$$

¹ マーチン・ガードナー (編), 田中勇 (訳), サム・ロイドのパズル百科, 1966 白揚社

6 Chapter.1 概要

となる。具体的数値を代入すると、

$$P[Y \geq 40] \approx 1 - \Phi(-1.656) \approx 0.95 = 95\%$$

となる。

Tea Time

..... ● 実験の計画が重要である：円周率の推定の場合

データを要約したら、データを支配する法則を推量したりする方法論の構築が統計学の使命である。しかし、統計的方法論の選択よりも、まず質の高いデータの取得がより重要であることを肝に銘じてほしい。ここで円周率 π を未知の定数として、 π を推定する問題を通して実験計画の重要性を考えよう。



図 1.1 同じ高さのマグカップ

半径 r 、高さ h の円柱状のマグカップの容積は

$$V_c = \text{底面積} \times \text{高さ} = (\pi r^2) \times h$$

となるので、円周率は

$$\pi = \frac{V_c}{hr^2} \quad (1.10)$$

となる。しかし、容積 V_c が分からないので、次のように間接的に V_c を測る²。

図 1.2 は、横幅 $a = 95\text{mm}$ 、縦幅 $b = 110\text{mm}$ 、一丁 400g 入れの豆腐容器である。マグカップにある満杯の水を豆腐容器にいれ、水の深さ $X(\text{mm})$ を測る。豆腐容器内の水の体積は

$$V_t = a \times b \times X \quad (\text{mm}^3)$$

であり、これがマグカップの容積 V_c の間接的な測定値と見なせる。

上の議論により、円周率の推定値として

$$\hat{\pi} = \frac{V_t}{hr^2} = \frac{ab}{hr^2} X \quad (1.11)$$

を得る。 X 以外を定数として、 $\hat{\pi}$ の精度を表す標準偏差は

$$\sqrt{V(\hat{\pi})} = \sqrt{\frac{a^2 b^2}{h^2 r^4} V(X)} \propto 1/r^2 \quad (1.12)$$

となり、マグカップの半径の 2 乗に応じて減少する。従って、円周率のよい推定量を得るため、

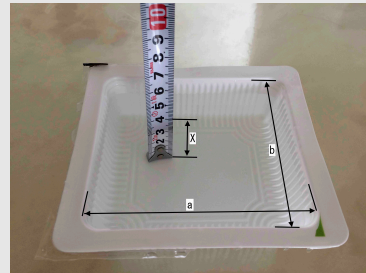


図 1.2 一丁 400g 入れの豆腐容器

できるだけ大口のマグカップを使うべきであることが分かる。表 1.1 は図 1.1 のマグカップを使った実験結果を纏めたものである。大口のカップを使った場合、一桁の精度向上が見られた。

マグカップ	半径 r	深さ h	水の深さ X	$\hat{\pi} = \frac{ab}{hr^2}X$
小口カップ	36	79	31	$\frac{323950}{102384} \approx 3.164$
大口カップ	41	79	40	$\frac{418000}{132799} \approx 3.148$

表 1.1 大小のマグカップによる円周率の推定値の比較

TEST shuffle 20

01

年

月

日

12

解く順序(問題の選択)

7

11

10

予想時間

(5)

(5)

(5)

(5)

(5)

(5)

(5)

実際の時間

(5)

(5)

(5)

(5)

(5)

(5)

(5

10

