目的

統計科学で最も重要な概念は**条件付確率** *(conditional probability)* である．
『社長と秘書』のたとえ話を使って，これをマスターしよう．

私のブログから

新しい製品のアイデアを思いついたあなたは社長に直接会って売り込みたい。できれば社長の機嫌が良いときに面会するのがよさそうだ。残念ながら、社長の機嫌が良いのか悪いのか、分からない。ところが社長の秘書とは毎日のように廊下ですれ違う。秘書の機嫌から社長の機嫌についての情報を読み取れるのではないか、とあなたは考えた。このとき、あなたは「条件付確率」を考えているのだ。

データサイエンスの中で最も重要な概念は条件付確率 (Conditional Probability) である。なぜなら、あることがら（秘書の機嫌）のことを知ることによって別のことがら（社長の機嫌）について推論することができればさまざまな応用が可能となるからである。秘書が機嫌が良いということが分かったとき、社長の機嫌が良いと期待できるだろう。逆に、秘書の機嫌が悪ければ面会はやめておいた方がよい。しかし、この重要な概念を正しく理解している学生は意外と少ない。

秘書の機嫌が良いことが分かった条件の下で、社長の機嫌が良い確率（条件付確率）が、秘書を見ずに社長の機嫌が良い確率（周辺確率）よりも大きければ、面会に行く意味がある。これらの確率の間にはどのような関係があるのだろう。また、社長も秘書もともに機嫌が良い、あるいは悪い確率（同時確率）とそれらの確率とはどう関係するのだろうか。さて、あなたの企みはそこそこうまくいき、社長とは飲み友達になった。ところが、恋心を感じ始めていた秘書とはほとんど会えなくなったしまった。そこで、社長の機嫌が良い日に秘書にデートを申し込むことを考えた。このとき、社長の機嫌が良いことが分かった条件の下で、秘書の機嫌が良い確率をあなたは考えていることになる。この逆の条件付確率と最初に書いた条件付確率の間にはどのような関係があるのだろうか。

私は講義やゼミの中で、最初は条件付確率にまったく不案内だった学生や社会人がこの「社長と秘書」の話を聞きながら、自分で具体的に確率を計算して考えることで、条件付確率の意味を正しく把握できるようになることを経験してきた。さらに、ベイズ推定や状態空間モデルを含む多くの応用、ひいてはモデルとシミュレーションの本質に話をつなげていくことができる。統計の教科書でこのあたりのことが強調されていないのはとても残念である。半日程度のセミナーでお話しすることができるので、興味を持たれた読者はぜひお声がけいただきたい。

自然現象であれ、社会現象であれ、現象のモデルは人がでっち上げる妄想であり、しかし時に極めて有用である。ただし、モデルには未知のパラメータや量が含まれるから、データからそれらを推論する必要がある。またモデルはただ一つではない。「秘書」がデータならば「社長」はモデルである。一方、モデルからシミュレーションを行ってデータを生成することはいつでも可能だ。したがって、秘書から社長の機嫌を推論することは「逆シミュレーション」であり、データ同化など別の名前でもよばれている。

話が長くなった。最後にクイズ。三つの扉がある。扉の向こうには、新車のポルシェが一つ、ハズレが二つ隠されている。あなたは扉の一つを選んだとする。このゲームの司会者が、あなたが選んでいない扉のうち、ハズレの扉を選んで開けてあなたに見せた。このとき、あなたは最初に選んだ扉のままで選択を変えないか、またはまだ開いていない扉を選びなおすか、どちらかをすることができると伝えられる。変えるべきか、変えないべきか。答えを知っている人は、次のバージョンの問題を考えられたい。司会者ではなく、風が吹いてたまたまあなたが選んでいない扉のうち一つが開いてしまい、それがハズレだった場合にはどうすべきか。

藤原 義久 / yoshi.fujiwara@gmail.com

『社長と秘書』：条件付確率とは何か

ある現象を直接に観測できないが，それと何らか関係がある別の現象から情報が得られることはよくある．例として，社長の機嫌と秘書の機嫌という二つの現象を考えよう．社長室に入れないが，秘書とは廊下でよくすれ違うので，秘書の機嫌から社長の機嫌を知りたいというわけだ．

人の機嫌とは不確かなもので**確率変数** *(stochastic varible)* としてあつかうのがよいだろう．社長の機嫌に対する確率変数を A で，秘書の機嫌に対する確率変数を B で表す．それらの取りうる値は以下の通りである．

事象 (event)	記号
社長の機嫌が良い	$A = a_1$
社長の機嫌が悪い	$A = a_2$

事象 (event)	記号
秘書の機嫌が良い	$B = b_1$
秘書の機嫌が悪い	$B = b_2$

機嫌は良いか悪いかのいずれかであり (排反的)，それらで可能な全事象となっている．観測した結果，以下のデータが得られたとしよう．

社長 \ 秘書	良い $B = b_1$	悪い $B = b_2$	合計
良い $A = a_1$	20	5	25
悪い $A = a_2$	15	60	75
合計	35	65	100

以下では観測された事象の数の割合として確率を計算する．

1. 社長の機嫌が良い確率と悪い確率をそれぞれ
$$P_A(a_i) := \mathrm{Prob}[A = a_i] \qquad (i = 1, 2) \tag{1}$$
と書こう．記号から明らかな場合，$P(a_i)$ のように略記する．社長の機嫌が良い確率 $P(a_1)$ と悪い確率 $P(a_2)$ はいくらか．

2. 秘書の機嫌が良い確率 $P(b_1)$ と悪い確率 $P(b_2)$ はいくらか[1]．

3. 社長の機嫌が良いまたは悪い，かつ秘書の機嫌が良いまたは悪いという**同時確率** *(joint probability)* を
$$P_{AB}(a_i, b_j) := \mathrm{Prob}[A = a_i \text{ and } B = b_j] \qquad (i, j = 1, 2) \tag{2}$$
と書く．これも $P(a_i, b_j)$ のように略記する．社長の機嫌が良く，かつ秘書の機嫌が良い確率 $P(a_1, b_1)$ はいくらか．

4. 社長の機嫌が良く，かつ秘書の機嫌が悪い同時確率 $P(a_1, b_2)$ はいくらか．

5. 次が成り立つことを示せ．
$$P(a_1) = P(a_1, b_1) + P(a_1, b_2) \tag{3}$$
上式が成り立つのは当然であることを理解せよ．$P(a_1)$ が上表の余白 (マージン) の部分から計算されるので，同時確率と区別するために $P(a_1)$ は**周辺確率** *(marginal probability)* とよばれる．

[1] $P_A(\cdot)$ と $P_B(\cdot)$ とは異なる関数であるので，同じ記号 $P(\cdot)$ を使うのはよくないが，引数の記号を見れば明らかである．同様に以下では引数の記号で区別することにする．

6. 秘書の機嫌が良いことが分かった条件の下で，社長の機嫌が良い確率，すなわち**条件付確率** *(conditional probability)* を

$$P_{A|B}(a_1|b_1) := \mathrm{Prob}[A = a_1 \text{ under the condition that } B = b_1] \tag{4}$$

と書く．記号から明らかな場合，$P(a_1|b_1)$ のように略記する．秘書の機嫌が良いことが分かった条件の下で，社長の機嫌が良い条件付確率 $P(a_1|b_1)$ を求めよ．

7. 条件付確率が

$$P(a_1|b_1) = \frac{P(a_1, b_1)}{P(b_1)} \tag{5}$$

を満たすことを確かめて，これが条件付確率の定義にふさわしいことを理解せよ．

8. 秘書の機嫌が悪いことが分かった条件の下で，社長の機嫌が良い条件付確率 $P(a_1|b_2)$ を求めよ．また

$$P(a_1|b_2) = \frac{P(a_1, b_2)}{P(b_2)} \tag{6}$$

を満たすことを確かめよ．

9. 以下の確率の大きさを比較せよ．

- 秘書の機嫌について情報がない中で，社長の機嫌が良い確率
- 秘書の機嫌が良いことが分かった条件の下で，社長の機嫌が良い確率
- 秘書の機嫌が悪いことが分かった条件の下で，社長の機嫌が良い確率

秘書の機嫌が良かろうと悪かろうと，その情報が分かれば社長の機嫌が良い可能性をよりよく推論できることを理解せよ．条件付確率という概念が重要であることの本質はこれである．

10. さて，社長とは縁もゆかりもない猫の機嫌を考えよう．その猫の機嫌を確率変数 C で，機嫌が良い・悪い場合をそれぞれ $C = c_1, C = c_2$ と表す．観測の結果，以下のデータが得られた．

社長 \ 猫	良い $C = c_1$	悪い $C = c_2$	合計
良い $A = a_1$	10	15	25
悪い $A = a_2$	30	45	75
合計	40	60	100

条件付確率 $P(a_1|c_1), P(a_1|c_2)$ を求めて

$$P(a_1|c_1) = P(a_1|c_2) \tag{7}$$

であることを確かめよ．猫の機嫌についての情報から社長の機嫌が分かるだろうか．

11. あたりまえの式 (3) と条件付確率の定義 (5),(6) から，もし (7) が成り立つならば

$$P(a_1, c_1) = P(a_1) \cdot P(c_1) \tag{8}$$

であることを示せ．逆に (8) が成り立てば，(7) も成り立つことも示せ．

このような場合，確率変数 A と C は**統計的に独立** *(statistically independent)* であるという．独立な確率変数については，一方の情報から他方の情報が得られない．

藤原 義久 / yoshi.fujiwara@gmail.com

12. ここまでは秘書の機嫌から社長の機嫌を知りたいと考えてきた．社長とは頻繁に会うようになったので，その機嫌は手に取るように分かるようになった．ところがなぜか秘書とはほとんど会えなくなった．そこで，逆に社長の機嫌から秘書の機嫌を知りたい．秘書にデートを申し込みたいわけだ．この場合，条件付確率として

$$P(b_1|a_1) = \frac{P(a_1, b_1)}{P(a_1)} \tag{9}$$

などを考えればよい．$P(b_1|a_1), P(b_1|a_2)$ を計算せよ．

- 社長の機嫌について情報がない中で，秘書の機嫌が良い確率
- 社長の機嫌が良いことが分かった条件の下で，秘書の機嫌が良い確率
- 社長の機嫌が悪いことが分かった条件の下で，秘書の機嫌が良い確率

の大きさを比較して，社長の機嫌についての情報から秘書の機嫌が良い可能性を推論できることを理解せよ．

13. (5) と (9) を連立して

$$P(a_i, b_j) = P(a_i|b_j)\,P(b_j) = P(b_j|a_i)\,P(a_i) \tag{10}$$

であることを示して，これから

$$P(b_j|a_i) = \frac{P(a_i|b_j)\,P(b_j)}{P(a_i)} \tag{11}$$

が得られることを確認せよ．これを**ベイズの定理** *(Bayes' theorem)* という.「定理」というほどのこともなく，条件付確率の定義からすぐに導けるものであることに注意する．

まとめ

一般に，確率変数 A があり (「社長」)，起こりうる事象は互いに排反的な (「機嫌は良いか悪いか」のどちらか) $A = a_i$ $(i = 1, 2, \ldots)$ であるとする．同様に確率変数 B も考えるとき

周辺確率	$P(a_i), \quad P(b_j)$
同時確率	$P(a_i, b_j)$
条件付確率	$P(a_i \vert b_j), \quad P(a_i \vert b_j)$

には以下の関係が成り立つ．これらは覚える必要はなく,『社長と秘書』の話からいずれも自明なものばかりであり，自分でいつでも導けることを強調しておく．

1. 全確率 (互いに排反な全事象について和をとると確率は 1)

$$\sum_i P(a_i) = 1, \qquad \sum_j P(b_j) = 1 \tag{12}$$

$$\sum_{i,j} P(a_i, b_j) = 1 \tag{13}$$

$$\sum_i P(a_i|b_j) = 1, \qquad \sum_j P(b_j|a_i) = 1, \tag{14}$$

2. 周辺確率と同時確率の関係[a]

$$P(a_i) = \sum_j P(a_i, b_j), \qquad P(b_j) = \sum_i P(a_i, b_j) \tag{15}$$

3. 条件付確率の定義

$$P(a_i, b_j) = P(a_i|b_j)\, P(b_j) = P(b_j|a_i)\, P(a_i) \tag{16}$$

4. ベイズの定理: (16) からすぐに導ける (ので，重要なのは条件付確率の定義とその意味である!)

$$P(b_j|a_i) = \frac{P(a_i|b_j)\, P(b_j)}{P(a_i)} = \frac{P(a_i, b_j)}{P(a_i)} \tag{17}$$

$$P(a_i|b_j) = \frac{P(b_j|a_i)\, P(a_i)}{P(b_j)} = \frac{P(a_i, b_j)}{P(b_j)} \tag{18}$$

[a]同時確率から周辺確率を得る操作は，周辺化 (marginalize) とよばれる．

藤原 義久 / yoshi.fujiwara@gmail.com

11 の解答: 統計的に独立な場合
条件付確率が条件によらないとき，すなわち $P(a_i|c_j)$ が j によらないとすると

$$\begin{aligned} P(a_i) &\overset{(15)}{=} \sum_j P(a_i, c_j) \\ &\overset{(16)}{=} \sum_j P(a_i|c_j)\, P(c_j) \\ &\overset{\text{仮定}}{=} P(a_i|c_j) \sum_j P(c_j) \\ &\overset{(12)}{=} P(a_i|c_j) \end{aligned} \tag{19}$$

これを (16) に代入すると

$$P(a_i, c_j) = P(a_i|c_j)\, P(c_j) = P(a_i) \cdot P(c_j) \tag{20}$$

逆に，(20) が成り立つとき

$$P(a_i|c_j) = \frac{P(a_i, c_j)}{P(c_j)} = \frac{P(a_i) \cdot P(c_j)}{P(c_j)} = P(a_i) \tag{21}$$

となり，$P(a_i|c_j)$ は j によらない.

参考文献

[1] 甘利 俊一「情報理論」(ダイヤモンド社，1970).
名著である．第一章に本稿に該当する箇所があるが，この章だけでも十分に読み応えがある.「社長と秘書」の話はここからとってふくらませたものである．ちくま学芸文庫から文庫本で最近出版されている.

練習問題

まれな病気と診断

まれな病気があり，人口の中の 0.1%だけがその病気を持っている．最新の診断技術によると，その病気にかかっている人が陽性である (病気であるという診断結果) と正しく診断される確率は 99%だが，実際には病気ではない人でも誤って陽性とされる確率が 2%あるという．

あなたが診断を受けた結果，残念ながら陽性であるということが分かったとき，本当にその病気にかかっている確率を知りたい．(ただし、あなたは人口の中でランダムに選ばれたものとし，病気に関係する他の情報は一切ないものとする)．

病気である・ない，陽性である・ないという組合せを次のような表にしてみよう．

	陽性である	陽性ではない
病気である	a	b
病気ではない	c	d

ここで

a 病気であり，かつ陽性である人の割合
b 病気であり，かつ陽性ではない人の割合
c 病気ではなく，かつ陽性である人の割合
d 病気ではなく，かつ陽性ではない人の割合

である．割合とは，人口全体に対する割合を意味する．したがって

$$a + b + c + d = 1 \tag{22}$$

上の題意を a, b, c, d を用いてすべて表現した上で，それらと式 (22) を用いて，陽性であるということが分かったとき，本当にその病気にかかっている確率を求めなさい (電卓などを使って計算してもよい)．

白黒カードの「マジック」

両面とも黒のカードが一枚，両面とも白のカードが一枚，片面が黒で片面が白のカードが一枚，合計で三枚のカードがある．それらのカードの順番も面もでたらめにシャッフルして，一枚だけ選んで机の上に置く．見えている面が黒であると分かったとき，その裏側の面が白であると賭けるのがよいか，黒であると賭けるのがよいか，あるいは白黒どちらと賭けても同じか．

クイズ: モンティ・ホール問題

3 つの扉がある。扉の向こうには、アタリの賞品が 1 つ，ハズレが 2 つ隠されている．あなたは扉の 1 つを選んだとする．このゲームの司会者は，あなたが選んでいない扉のうち，ハズレの扉を選んで開けてあなたに見せてくれる．このとき，あなたは最初に選んだ扉のままで選択を変えないか，またはまだ開いていない扉を選びなおすか，どちらかを選択することができると伝えられる．変えるべきか、変えないべきか．

司会者ではなく，風が吹いてたまたまあなたが選んでいない扉のうち 1 つが開いてしまい，それがハズレだった場合にはどうすべきか．