情報システム工学科
伊達 章

2005 年 4 月 11 日

速修 確率統計

医者の藪野竹庵氏は自分のところへくる患者について

$$A : \begin{cases} A_1 : & \text{熱がある} \\ A_2 : & \text{熱がない} \end{cases} \qquad B : \begin{cases} B_1 : & \text{風邪を引いている} \\ B_2 : & \text{風邪ではない} \end{cases}$$

の 2 種類の事象の割合を調べた．これらの事象を組み合わせると，4 通りの複合事象ができる．事象 (A_i, B_j) の起こる割合は次表のようであった．この割合は，事象 (A_i, B_j) の確率 $p(A_i, B_j)$ であると思ってさしつかえない．

	B_1（風邪）	B_2（風邪なし）	$p(A_i)$
A_1（熱あり）	0.55	0.05	0.60
A_2（熱なし）	0.10	0.30	0.40
$p(B_j)$	0.65	0.35	

この表からわかるように，熱があるかないかを知ってしまえば，風邪であるかないかはほぼ検討がつく．必要な確率の知識を整理しておこう．確率分布として次のものを考える．

$$\begin{cases} \text{同時確率分布:} & p(A_i, B_j) \\ \text{周辺確率分布:} & p(A_i), p(B_j) \\ \text{条件付確率分布:} & p(B_j|A_i), p(A_i|B_j) \end{cases}$$

$p(A_i, B_j)$ のように，A_i と B_j とが組になって同時に起こる確率を示すものを，同時確率分布と呼ぶ．そのうちの一方だけに着目した確率分布 $p(A_i)$ や $p(B_i)$ は，上の表を行または列について加え合わせると出てくる．この $p(A_i)$ や $p(B_i)$ は，表の周辺に並んでいるので，周辺分布と呼ぶ．条件付確率分布 $p(B_j|A_i)$ とは A に関する事象が A_i であったときの B_j の確率である．$p(A_i|B_j)$ は逆に B が B_j であることがわかっているときの A_i の確率である．これらの間には次の関係が成立する．

$$\sum_{i,j} p(A_i, B_j) = 1,\ \sum_i p(A_i) = \sum_j p(B_j) = 1 \tag{1}$$

$$\sum_j p(B_j|A_i) = \sum_i p(A_i|B_j) = 1 \tag{2}$$

$$p(A_i) = \sum_j p(A_i, B_j),\ p(B_j) = \sum_i p(A_i, B_j) \tag{3}$$

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

$$p(B_j|A_i) = \frac{P(B_j)P(A_i|B_j)}{p(A_i)} = \frac{P(A_i, B_j)}{p(A_i)} \tag{5}$$

$$p(A_i|B_j) = \frac{P(A_i)P(B_j|A_i)}{p(B_j)} = \frac{P(A_i, B_j)}{p(B_j)} \tag{6}$$

式 (4) は，A_i と B_j との同時確率は，A_i の確率に A_i が起こるという条件下での B_j の起こる確率を掛けたものであることを示す．式 (5,6) は Bayes の公式と呼ばれるもので，式 (4) から導き出せる．A と B が確率的に独立な場合（風邪と熱が無関係の場合）には

$$p(A_i, B_j) = p(A_i)p(B_j) \tag{7}$$

$$p(B_j|A_i) = p(B_j), p(A_i|B_j) = p(A_i), \tag{8}$$

という式が成立する．