

ベイズ理論 Bayesian Theory

事象 A と B が有る時、「A 且つ B」が生じる確率は、「A が生じる確率」 $p(A)$ × 「A が生じたと言う前提の下で B が生じる条件付確率」 $p(B|A)$ と表せるが、同様に $p(B) \cdot p(A|B)$ とも表され、この二つは同じ確率を意味する。

$$p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

因みに、「A と B が独立事象である」と云うのは、条件付確率が条件なしの確率に等しい事として定義されているので、

$$p(B|A) = p(B), p(A|B) = p(A)$$

よって、

$$p(A) \cdot p(B|A) = p(B) \cdot p(A|B) = p(A) \cdot p(B)$$

となる。

架空例であるが、ある地域で、ドライバーの人口比が男性 60% (.60)、女性 40% (.40)、男性ドライバーが交通事故を起こす確率が 10% (.10)、女性が交通事故を起こす確率が 8% (.08) だとする。

「ドライバーが男性である」と「ドライバーが事故を起こす」の確率は、

$$p(\text{male}) \cdot p(\text{accident}|\text{male}) = .60 \cdot .10 = .06 = p(\text{accident}) \cdot p(\text{male}|\text{accident})$$

ここで、

$$p(\text{accident}) = p(\text{male}) \cdot p(\text{accident}|\text{male}) + p(\text{female}) \cdot p(\text{accident}|\text{female}) = .60 \cdot .10 + .40 \cdot .08 = .092$$

であるので、次の式が得られる。

$$p(\text{male}|\text{accident}) = .60 \cdot .10 / .092 = .6521739$$

これは、「事故が起こった時に、そのドライバーが男性である確率」を意味している。

A を「或証拠 e が得られる」事とし、B を「或仮説 H_0 が正しい」事とする。

$$p(e) \cdot p(H_0|e) = p(H_0) \cdot p(e|H_0)$$

証拠 e は、 H_0 が正しくて e が得られる場合と、 H_0 が誤っていて(他の仮説が正しくて) e が得られる場合からなるので、

$$p(e) = p(H_0) \cdot p(e|H_0) + p(\neg H_0) \cdot p(e|\neg H_0) = p(H_0) \cdot p(e|H_0) + p(H_1) \cdot p(e|H_1) + p(H_2) \cdot p(e|H_2) + \dots$$

これらから、

$$\begin{aligned} p(H_0|e) &= \frac{p(H_0) \cdot p(e|H_0)}{p(e)} = \frac{p(H_0) \cdot p(e|H_0)}{p(H_0) \cdot p(e|H_0) + p(\neg H_0) \cdot p(e|\neg H_0)} \\ &= \frac{p(H_0) \cdot p(e|H_0)}{p(H_0) \cdot p(e|H_0) + p(H_1) \cdot p(e|H_1) + p(H_2) \cdot p(e|H_2) + \dots} \end{aligned}$$

となる。この確率は、「証拠 e が得られた時に、仮説 H_0 が正しい」事後確率と呼ばれ、「仮説 H_0 が正しい」事前確率と、「仮説 H_0 が正しい時に、証拠 e が得られる」条件付確率(尤度とも呼ばれる)などから計算される。

「或理論(仮説)が正しいならば或証拠が得られる確率」と「或理論(仮説)が正しくなくて或証拠が得られる確率」が分かっている時、「或理論(仮説)が正しい」事前確率を与えれば、「その証拠が得られた場合に、その理論が正しい」事後確率を求められるのである。

例えば、或理論が正しいければ特定の観察結果が得られる確率が 50%、その理論が正しくなくてその観察結果が得られる確率が 20% だとする。当該の観察結果が得られたと云う結果から、その理論が正しい事後確率を求める為には、その理論が正しい事前確率を決める必要がある。全く何の情報も無く「正しいかどうかさっぱり分からない」と云う場合には、「五分五分」「フィフティ・フィフティ」の言葉通り、事前確率に 50% を与える事が多い。さて、事後確率はどうなるだろうか? これを、データによる確率(信念)の更改 *update* と呼ぶ。

従来の頻度主義統計学の地位を大きく脅かしているベイズ統計学であるが、この様にベイズ理論は、何らかの方法で条件付確率を適切に数値化する事が出来れば、単一の観察・事例によっても理論の検討が出来るので、歴史学や政治学などの定性的(質的)研究でも注目を集めている。