

前回の補足：連続型確率変数の例

高いところから針を落とし、座標を記録する作業を何度も繰り返しその x 座標を考える。このとき針の太さは無いとすると、どの x 座標についてもその座標に正確に針が落ちることは有りそうに無い。しかし針を落とした全回数を分母として座標が a 以下になった回数の相対的な頻度を考えることは可能である。

x 座標を確率変数 X でモデル化することができる。このとき累積分布関数 $F_X(a)$ は上記の座標が a 以下になる相対的な頻度に対応する。また、確率関数 P_X はどの値に対しても 0 になるが、そのことはどの座標についても丁度そこに針が落ちる事象の相対的な頻度は 0 であることに対応する。

確率は、事象が起こる相対的な頻度であると解釈できる場合もある。

今日の内容：多次元の確率分布パート 1

(教科書の 7 章)

同時確率分布（または結合確率分布）

確率論を現実を使用するときに、確率変数で表わされる現象は一つではなく複数あることが多い。したがって、そのような現象を記述するときには、複数の確率変数を同時に扱う必要がある。今日の講義では、複数の確率変数を扱うための道具を紹介する。

同時累積分布関数

確率変数 X, Y があるとする。 X が x 以下の実現値をとり**かつ** Y が y 以下の実現値をとる確率を

$$P(X \leq x, Y \leq y) = F_{XY}(x, y)$$

で表わし、 F_{XY} を同時累積分布関数と呼ぶ。

一般に $F_{XY}(x, y) \neq F_X(x) \cdot F_Y(y)$ であるから注意。

同時累積分布関数の例

今までの講義で例として用いた赤玉と黄玉の例を考え、 X を1回目に赤玉が出たときに1 それ以外は0, Y を2回目に赤玉が出たときに1 それ以外は0, Z を1回目と2回目に赤玉が出た回数の合計とする。

このとき

$$F_{XY}(x, y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ 0.25 & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y < 1 \\ 0.5 & \text{if } 0 \leq x < 1 \text{ and } 1 \leq y \\ 0.5 & \text{if } 1 \leq x < 2 \text{ and } 0 \leq y < 1 \\ 1 & \text{if } 1 \leq x < 2 \text{ and } 1 \leq y \end{cases}$$

となる。ここで $F_{XY} = F_X(x)F_Y(y)$ が成立していることに注意 (F_X, F_Y を図示する)

同時確率質量関数または同時確率分布

確率変数 X が x になりかつ Y が y になる確率を

$$P(X = x, Y = y) = P_{XY}(x, y)$$

で表わし、 P_{XY} を同時確率質量関数または同時確率分布と呼ぶ。一般に $P_{XY}(x, y) \neq P_X(x) \times P_Y(y)$ である。

例: X, Y, Z を前ページで定めた通りとする。 P_{XY} の値は以下の表のようになる。

$y \setminus x$	0	1
0	0.25	0.25
1	0.25	0.25

このときすべての x, y について $P_{XY}(x, y) = P_X(x) \times P_Y(y)$ である。

一方 X と Z の同時確率質量関数 P_{XZ} の値は以下のようになる。

$z \setminus x$	0	1
0	0.25	0
1	0.25	0.25
2	0	0.25

このとき $P_Z(0) = P_Z(2) = 0.25$, $P_Z(1) = 0.5$ なので

$$0 = P_{XZ}(0, 2) \neq P_X(0)P_Z(2) = 0.125$$

である。

同時確率密度関数

確率変数 X, Y がそれぞれ連続型であるとする。もし

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(x, y) dx dy$$

となる関数 f_{XY} が存在する場合、 f_{XY} を同時確率密度関数と呼ぶ。

同時確率密度関数 f_{XY} は必ずしも存在するとは限らないが、同時累積分布関数 F_{XY} は常に存在する。また f_{XY} が存在する場合 f_{XY} は F_{XY} から計算できる。しかし、実用上は同時確率密度関数または同時確率質量関数で (X, Y) の確率的な振舞を完全に記述できる場合が大半なので、今後 f_{XY} または P_{XY} が与えられているものとして説明する。

条件付きの確率分布

複数の確率変数を扱っているとき、一部の確率変数の値が観測などでわかり残りの確率変数の値が未知の場合がある。そのようなときに未知の確率変数の確率分布を考えるための道具を導入する。

X, Y が離散型で $Y = y$ という条件の下での X の確率質量関数を

$$P_{X|Y}(x|y) = \frac{P_{XY}(x, y)}{P_Y(y)}$$

で定義し、これを条件付き確率 (質量) 関数または条件付き確率分布と呼ぶ。

例: X, Y, Z を赤玉、黄玉の例とすると $P_{Z|X}(0|0) = P_{Z|X}(1|0) = 0.5$, $P_{Z|X}(2|0) = 0$ で、1 回目に黄玉を取り出すと、赤玉を取り出した回数の合計は 2 に成り得ないことを示している。

X, Y が連続型確率変数のときは条件付き確率密度関数

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

を代わりに考える。

周辺確率分布

2つの確率変数 X, Y があるときに X の単独の分布 P_X は Y の値が何であるか知らないときの X の分布を表わしていると考えられる。 P_X を P_{XY} から計算するためには

$$P_X(x) = \sum_y P_{XY}(x, y)$$

とすればよい。但し、総和記号は Y の取り得る値すべてについて和を取っている。こうすることにより Y の値はなんでもよいが $X = x$ となる確率を考えることになり $P_X(x)$ が得られる。 P_X を P_{XY} の周辺確率(質量)関数または周辺分布と呼ぶ。

X, Y が連続型のときは

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

という計算で、同時確率密度関数 f_{XY} から X だけの確率密度関数 f_X が得られる。 f_X を f_{XY} の周辺確率密度関数と呼ぶ。

確率変数の独立性

X, Y が離散型確率変数とする。すべての x, y について

$$P_{XY}(x, y) = P_X(x)P_Y(y)$$

が成立するとき X と Y は統計的に独立であると言う。とくに $P_Y(y) \neq 0$ である y について

$$P_{X|Y}(x|y) = \frac{P_{XY}(x, y)}{P_Y(y)} = P_X(x)$$

が成立し、このことは確率変数 Y の値を観測しても条件付き確率分布 $P_{X|Y}$ は y の値を知らないときの X の確率分布 P_X と変わらないことを示している。

例：赤玉と黄玉の例では X と Y は統計的に独立で1回目に赤玉を取り出すことと2回目に赤玉を取り出すことは関係が無い。一方 X と Z は統計的に独立ではなく、とくに1回目に赤玉を取り出したかどうかは2回の取り出しで赤玉を取り出した回数の合計に影響がある。

X, Y が連続型確率変数の場合、すべての x, y について

$$f_{XY}(x, y) = f_X(x)f_Y(x)$$

が成立するときに、 X と Y は統計的に独立であると言う。

確率変数の積の期待値

確率変数 X, Y があり、それらの積 Z の期待値を求めたいとする。この場合 Z の確率質量関数 P_Z または確率密度関数 f_Z を計算する必要があるように思われるが、実は不必要である。 X, Y が離散型の場合

$$E[Z] = \sum_z P_Z(z)z = \sum_{x,y} P_{XY}(x,y)xy$$

が成立する。同様に X, Y が連続型の場合

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x,y) dx dy$$

が成立する。

一般には $E[XY] \neq E[X]E[Y]$ であるが、 X と Y が統計的に独立ならば $E[XY] = E[X]E[Y]$ が成り立つ。理由は

$$\begin{aligned} E[XY] &= \sum_{x,y} P_{XY}(x,y)xy \\ &= \sum_{x,y} P_X(x)P_Y(y)xy \\ &= \sum_{x,y} P_X(x)x \cdot P_Y(y)y \\ &= \left(\sum_x P_X(x)x \right) \left(\sum_y P_Y(y)y \right) \\ &= E[X]E[Y] \end{aligned}$$

X, Y が連続型である場合も同様。

次回予告：多次元の確率分布パート 2 (教科書の 7 章)

演習問題

原則として授業終了までに答案を提出して下さい。終らない場合は授業が行われた日の夜までに南3号館311号室前の箱に演習答案を提出して下さい。

21. 正四面体に0,1,2,3の数字が書いてある4面サイコロを考える。このサイコロを振ると各数字が等しい確率で出るとする(床に面している数字が「出た」と見なす)。このサイコロの出た目を確率変数 X で表わし、 X を3で割った余りを $Y (= 0,1,2)$ とする。 X, Y の同時確率分布を求めよ。

22. 上記の場合に条件付き確率分布 $P_{X|Y}$ を求めよ。

23. X と Y は統計的に独立か否か理由と共に答えよ。

24. $E[XY]$ と $E[X]E[Y]$ をそれぞれ求めよ。

25. 確率変数 Z を X を2で割った余り ($= 0,1$) とする。 X と Z は統計的に独立か否か理由と共に答えよ。

ここからは以下のような累積分布関数を持つ確率変数 X を考える。

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{3}{4}x^2 & \text{if } 0 \leq x \leq 1, \\ -\frac{1}{4}(x-2)^2 + 1 & \text{if } 1 \leq x \leq 2, \\ 1 & \text{if } 2 \leq x \end{cases}$$

26. X の期待値、最頻値、中央値、分散、標準偏差を求めよ。
27. X の標準化変数 Z の累積分布関数を求めよ。
28. Z の確率密度関数を図示せよ。
29. 「期待値」、「分散」、「確率変数の独立性」、「仮説検定」を高校で習ったかどうか書いて下さい。授業でわかりにくかった所や要望を書いて下さい。また前回の答案で×になったところで、理由がわからないところは申し出て下さい。