確率への招待　13回目

確率変数と確率分布①

１．確率変数と確率分布

これまで学んできた確率の考えを、もっと数学的に扱いやすくしたい。

・「サイコロを振ったときに出た目」は数字だから、まぁそのままでも数学的に扱える。

・「試合で勝った」「負けた」は、このままでは数字ではないから、扱いづらい

⇒ 試合に勝った・・・１
　　〃　負けた・・・０

と置き換えれば、数学的に扱いやすい。

これが「確率変数」。

・つまり、確率変数とは起こった事象を数字に置き換える変数のこと。

例）１個のサイコロを振るとき、出た目の値をXとすると、
Xのとり得る値は１，２，３，４，５，６の６つで、
各々の確率は1/6である。

別の変数Yを、

$$
Y = \begin{cases} 1 & （出た目が１か３のとき） \\ 0 & （出た目が2，4，6のとき） \\ -1 & （出た目が5のとき） \end{cases}
$$

とすると、Y＝１となる確率は1/3、0となる確率は1/2、
－１となる確率は1/6である。

このように、確率変数は事象と対応しているので、各値をとる確率が定まっている。

一般に、取りうる値とその確率が与えられた変数のことを確率変数という。

定義だけ見るとヤヤコシソウだが、別に大したことをやっているわけではない。

「お昼にカレーを食べた」「定食を食べた」などと言葉で書いていると数学的に扱いにくいので、数字に置き換えた、ということ。

例えば、

・家庭の電力使用量・・・もともと数字なので、このまま確率変数として考えてもよい

・犯罪捜査・・・犯人ならば1、そうでなければ0をとる確率変数

・アンケートの結果・・・ある政策に賛成なら1、反対なら2、どちらでもないなら9

など。

確率変数Xのとり得る値がx_1、x_2、…、x_nであるとき、Xが一つの値x_kをとる確率をP(X＝x_k)で表す。

前々ページのサイコロの例でいうと、

$$P(X=1)=\frac{1}{6}, P(X=2)=\frac{1}{6}, \cdots, P(X=6)=\frac{1}{6}$$

$$P(Y=-1)=\frac{1}{6}, P(Y=0)=\frac{1}{2}, P(Y=1)=\frac{1}{3}$$

これを次のような表にしておくと見やすい。

X	1	2	3	4	5	6	計
P	1/6	1/6	1/6	1/6	1/6	1/6	1

Y	－1	0	1	計
P	1/6	1/2	1/3	1

この対応関係を**確率分布**（または単に**分布**）といい、確率変数はこの分布に**従う**という。

確率なので、**下の欄の数字は0から1の値をとり、合計は1**。

２．確率変数の期待値と分散

確率変数Xが概ねどのような値をとるかを考えてみる。
たとえば、サイコロの目のように全ての値を取りうる値が同様に確からしいときには、その期待値（平均）を

$$\frac{1+2+\cdots+6}{6}=3.5$$

のように算術平均で考える。

では、同様に確からしくないときはどうだろうか。
サイコロの目で1の目を2の目に変えたとする。このとき、2の目の出る確率は1/3となり、この場合の期待値は

$$\frac{2+3+3+4+5+6}{6}=\frac{1}{3}\times 2+\frac{1}{6}\times 3+\cdots+\frac{1}{6}\times 6=3.67$$

となる。

つまり、取りうる値とその確率を掛けて足し合わせることで期待値を計算できる。

確率変数Xが以下のような分布に従うとする。

X	x_1	x_2	・・・	x_n	計
$P(X=x_k)$	p_1	p_2	・・・	p_n	1

このとき、Xの期待値または平均を

$$x_1 p_1 + x_2 p_2 + \cdots + x_n p_n = \sum_{k=1}^{n} x_k p_k$$

で定義し、記号E(X)またはmで表す。
（Eはexpectationの、mはmeanの頭文字）

中学校で度数分布表（ヒストグラム）を使って平均値を求めたのとまったく同じ。

これは「全体をおしなべると、Xがだいたいどのくらいの値であるか」を表すもの。

次に、「確率変数Xがどの程度ばらついているか」を考える。
x_k が平均 m からどれくらい離れているかは、引き算してx_k-mと計算されるが、$\Sigma(x_k-m)p_k=0$である。

$$\sum(x_k-m)p_k=\sum x_k p_k-m\sum p_k=m-m\cdot 1=0$$

・絶対値をとって$|x_k-m|$とすると、$\Sigma|x_k-m|p_k$はばらつきを表す指標となる（平均絶対偏差という）が、絶対値が数学的に扱いにくい（微分不可能）こともあり、あまり使われない。

・2乗して$\Sigma(x_k-m)^2p_k$を考えると、これは正の数となり、数学的にも扱いやすい（例えば、最大や最小を求めるとき、微分すると1次関数になって簡単に解ける）ので、これがよく用いられる。
これを確率変数の分散といい、記号V(X)で表す。

$$V(X)=\sum_{k=1}^{n}(x_k-m)^2p_k=E((X-m)^2)$$

Ｖ(X)の「単位」はxの2乗の単位になってしまう（例えば、xが長さ（メートル）ならば、Ｖ(X)は平方メートル）ので、やや不便。

⇒Ｖ(X)の平方根（$\sqrt{V(X)}$）はXと同じ単位になるので、

これをXの**標準偏差**といい、記号σ(X)で表す。

（「しぐま　エックス」、σはstandard deviationの頭文字）

ついでにもう一つ。

$$V(X) = \sum_{k=1}^{n}(x_k - m)^2 p_k = \sum_{k=1}^{n}(x_k^2 - 2x_k m + m^2)p_k$$

$$= \sum_{k=1}^{n} x_k^2 p_k - 2m \sum_{k=1}^{n} x_k p_k + m^2 \sum_{k=1}^{n} p_k$$

$$= E(X^2) - 2m \times m + m^2 \times 1 = E(X^2) - m^2 = E(X^2) - (E(X))^2$$

この、Ｖ(X)＝Ｅ(X^2)－(Ｅ(X))^2　という結果は、あとでも使うので、覚えておいた方がよい。

平均と分散の計算方法

p.5の確率変数XとYの平均と分散を計算してみる。

確率変数の値とその確率を掛けたものを計算し、その合計が平均となる。また、確率変数の値の2乗と確率を掛けたものを計算し、その合計は確率変数の2乗の平均となる。
この結果に基づいて、分散は$V(X) = E(X^2) - (E(X))^2$となる。

X	1	2	3	4	5	6	計	
P	1/6	1/6	1/6	1/6	1/6	1/6	1	
XP	1/6	2/6	3/6	4/6	5/6	6/6	7/2	(＝$E(X)$)
X²P	1/6	4/6	9/6	16/6	25/6	36/6	91/6	(＝$E(X^2)$)

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Y	－１	０	１	計	
P	1/6	1/2	1/3	1	
YP	–1/6	0	1/3	1/6	(＝$E(Y)$)
Y²P	1/6	0	1/3	1/2	(＝$E(Y^2)$)

$$V(Y) = \frac{1}{2} - \left(\frac{1}{6}\right)^2 = \frac{17}{36}$$

期待値の(一つの)特徴づけ

　aを定数とするとき、$E((X-a)^2)$をaの回りの平均2乗偏差と呼ぶが、これが最小になるのは、$a=E(X)$のときである。

(証明)

$(X-a)^2=\{(X-E(X))+(E(X)-a)\}^2$

$=(X-E(X))^2+2(X-E(X))(E(X)-a)+(E(X)-a)^2$なので、

これの期待値をとると、

$E((X-a)^2)=E((X-E(X))^2)+(E(X)-a)^2=V(X)+(E(X)-a)^2$

よって、aを変数と考えてこれが最小となるのは、

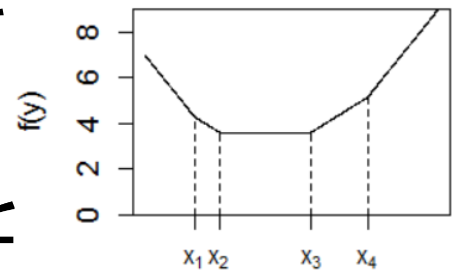
$a=E(X)$のとき。

それでは、平均絶対偏差だとどうなるだろう？

確率変数Xの取り得る値x_1, x_2, …, x_nが全て等確率とする。aのまわりの平均絶対偏差E(|X－a|)が最小になるのは、aがXの中央値のときである。

（証明）

Xの取りうる値を小さい方から順に並べて
$x_1 \leqq x_2 \leqq \cdots \leqq x_n$とする。
関数$f(y) = \sum \{|x_i - y|\}$のグラフを考えると



$$f(y) = \begin{cases} -ny + (x_1 + x_2 + x_3 + \cdots) & y \leq x_1 \\ -(n-2)y + (-x_1 + x_2 + x_3 + \cdots) & x_1 < y \leq x_2 \\ \vdots & \vdots \\ ny - (x_1 + x_2 + x_3 + \cdots) & x_n < y \end{cases}$$

と折れ線のグラフになり、これが最小値をとるのは、
nが偶数＝2kならば、$x_k \leqq y \leqq x_{k+1}$のとき
（最小値をとるyはたくさんある）
nが奇数＝2k＋1ならば、$y = x_{k+1}$のとき。

３．確率変数の変換

次のような分布に従う確率変数Xに対し、

X	x_1	x_2	・・・	x_n	計
$P(X=x_k)$	p_1	p_2	・・・	p_n	1

Xの一次関数Y＝aX＋b（ただしa，bは定数）も確率変数。

Y	y_1	y_2	・・・	y_n	計
$P(Y=y_k)$	p_1	p_2	・・・	p_n	1

$(y_i=ax_i+b)$

Xに対してこのようなYを考えることを確率変数の変換という。
（一次関数ならば一次変換、または線形変換という）

このとき、Yの期待値や分散、標準偏差がどうなるか考える。

$$E(Y)=\sum y_k p_k=\sum(ax_k+b)p_k=a\sum x_k p_k+b\sum p_k$$
$$=aE(X)+b$$

「確率変数の一次変換と、期待値をとる操作とは、順序の入れ替えが可能」

次に分散を計算してみると、

$y_k - E(Y) = ax_k + b - \{aE(X) + b\} = a\{x_k - E(X)\}$だから、

$$V(Y) = \sum_{k=1}^{n} \{y_k - E(Y)\}^2 p_k = a^2 \sum_{k=1}^{n} \{x_k - E(X)\}^2 p_k = a^2 V(X)$$

よって、標準偏差は、

$$\sigma(Y) = \sqrt{V(Y)} = \sqrt{a^2 V(X)} = |a| \sigma(X)$$

まとめると、

$$E(aX + b) = aE(X) + b,$$

$$V(aX + b) = a^2 V(X),$$

$$\sigma(aX + b) = |a| \sigma(X)$$

ところで、「一次変換と期待値をとる操作とは順序交換可能」だったが、一次変換以外だと、こううまくはいかない。

例）Y＝X^{2}という確率変数の変換に対して期待値をとると、

$E(X^2) = \{E(X)\}^2 + V(X)$ である（p.9でやった）。

よって、$V(X) = 0$ でない限り、$E(X^2) > \{E(X)\}^2$

一般には次のことが成り立つ。

Jensenの定理）下に凸の関数f(X)に対し、E(f(X))≧f(E(X))

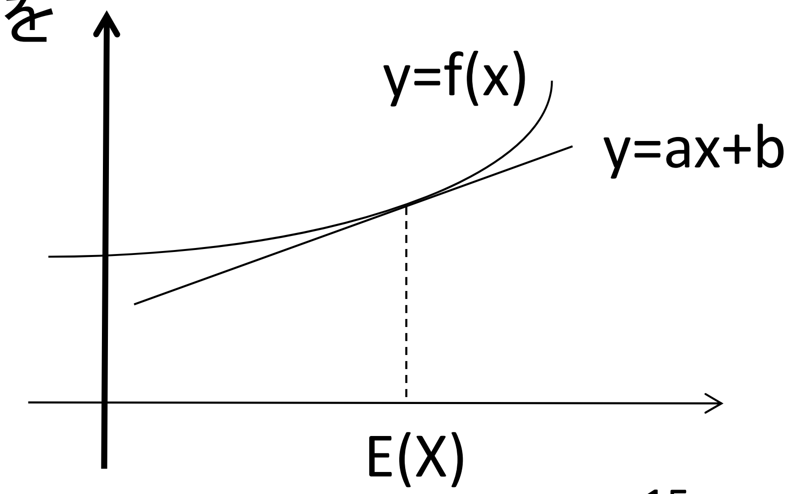
証明）f(X)が下に凸なので、右図のように、

x=E(X)のところでy=f(x)に接線y=ax+bを引くと、y=f(x)はこの接線の上にくる。

すなわち、任意のxに対しf(x)≧ax+b

両辺の期待値をとって、

E(f(X))≧E(aX+b)=aE(X)+b=f(E(X))



４．２つの確率変数の同時分布、確率変数の独立

（１）同時分布

　X、Yを確率変数とするとき、実数a，bに対して

　　X＝a　かつ　Y＝b　となる確率を　P（X＝a，Y＝b）と表す。

一つの確率変数について$X=x_k$となる確率を表で表したのと同様、(X,Y)についても、P（$X=x_i$，$Y=y_j$）＝p_{ij}となるp_{ij}を行列の形で書き表すと、

	y_1 y_2 …… y_m	計
x_1	p_{11} p_{12} … p_{1m}	p_1
x_2	p_{21} p_{22} … p_{2m}	p_2
:		
:		
x_n	p_{n1} p_{n2} … p_{nm}	p_n
計	q_1 q_2 … q_m	1

となる。この対応をXとYの同時分布という。

また、この表から、

$$P(X = x_i) = \sum_{j=1}^{m} p_{ij} = p_i,\quad P\left(Y = y_j\right) = \sum_{i=1}^{n} p_{ij} = q_i$$

となる。したがって、表の右端の欄は確率変数Xの確率分布、表の一番下の欄は確率変数Yの確率分布となっている。

これをそれぞれ、Xの周辺分布、Yの周辺分布という。

（2）確率変数の独立

２つの確率変数X、YについてXとYが独立であるということを、事象の独立と同様に、次のように定義する。

任意の実数a，bに対し、P（X＝a，Y＝b）＝P（X＝a）P（Y＝b）となるとき、XとYは独立であるという。

先ほどの同時分布の表でいうと、
「任意のi，jについて、$p_{ij}=p_i \times q_j$ が成り立つとき、XとYは独立である」ということになる。

先週まででやった「事象の独立」では、P（A∩B）＝P（A）P（B）となるとき、事象Aと事象Bは独立であるというのであった。
これと同じことである。