統計的な推測【確率変数の期待値と分散】 p.56〜60

青チャートB: 例題 63〜67

【態度目標】しゃべる、質問する、説明する、動く、協力する、貢献する

【内容目標】確率変数の期待値，分散，標準偏差を求めることができるようになろう

□確率変数の期待値

	賞 金	本 数
1 等	10000 円	5 本
2 等	1000 円	20 本
3 等	200 円	75 本
はずれ	0 円	900 本
計		1000 本

例題 1 の 1000 本のくじにおいて，賞金額の平均は，確率変数 X の値と，その値をとる確率を用いて，次のように求められる。このくじ 1000 本の賞金の総額は

$$10000\cdot5+1000\cdot20+200\cdot75+0\cdot900=85000$$

である。これを，くじの総数で割ると

$$\frac{1}{1000}(10000\cdot5+1000\cdot20+200\cdot75+0\cdot900)=85 \quad \cdots\cdots ①$$

となる。この値は，くじ 1 本あたりの賞金額の平均である。

すなわち，くじを 1 本引くときに期待できる賞金額は 85 円と考えられる。

このことは，次のように考えることもできる。

このくじを 1 本引くときに得る賞金を X 円とすると，確率変数 X の確率分布は次の表のようになる。

X	10000	1000	200	0	計
P	$\frac{5}{1000}$	$\frac{20}{1000}$	$\frac{75}{1000}$	$\frac{900}{1000}$	1

縦に掛けて横に足す

ここで，等式 ① を分配して整理すると，次のように書き表すこともできる。

$$10000\cdot\frac{5}{1000}+1000\cdot\frac{20}{1000}+200\cdot\frac{75}{1000}+0\cdot\frac{900}{1000}=85$$

この場合，くじを 1 本引く試行の結果得られる賞金額は，X の値として 4 通りの場合があるが，平均的に期待される賞金額は 85 円である。したがって，この等式の左辺は，賞金の額とそれが当たる確率の積をすべて加えたものになっていることがわかる。

確率変数 X が下の表に示された分布に従うとする。

X	x_1	x_2	……	x_n	計
P	p_1	p_2	……	p_n	1

縦に掛けて横に足す

期待値は数学Aでも出てきたが $E(X)$ や Σ を用いてはいなかった。

このとき

$$x_1p_1+x_2p_2+\cdots\cdots+x_np_n=\sum_{k=1}^{n}x_kp_k=\sum_{k=1}^{n}x_kP(X=x_k)$$

を，X の **期待値** または **平均** といい，$\boldsymbol{E(X)}$ または $\boldsymbol{m}$ で表す。

教科書では，分布に対しては「平均」，確率変数については「期待値」を用いている

統計的な推測【確率変数の期待値と分散】 p.56～60

青チャートB: 例題 63～67

確率変数の期待値

$$E(X)=x_1p_1+x_2p_2+\cdots\cdots+x_np_n=\sum_{k=1}^{n}x_kp_k$$

期待値（平均）は、「確率分布の中心を示すもの」

$E(X)$ の E は，
期待値を意味する英語
expectation の頭文字である。
m は平均を意味する英語
mean の頭文字である。

補足 和の記号 Σ

数列 $\{a_n\}$ について，初項から第 n 項までの和を，記号 $\sum$ を用いて $\sum\limits_{k=1}^{n}a_k$ と書く。

$$\sum_{k=1}^{n}a_k=a_1+a_2+a_3+\cdots\cdots+a_n$$

また，$\sum\limits_{k=p}^{q}a_k$ と書けば，数列 $\{a_n\}$ の第 p 項から第 q 項までの和を表す。

(*) Σ は，英語の sum (和) の頭文字 S に対応するギリシャ文字で，シグマと読む。

p_k に規則性がなければ
Σ でまとめて書く必要はない。
また，Σ の公式を知らなければ
Σ を活用する必要はない。

例 1） 1 枚の硬貨を 2 回続けて投げる試行で，表の出る回数を X とすると，X の確率分布は右の表のようになる。

したがって，X の期待値は $E(X)$ は

X	0	1	2	計
P	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

縦に掛けて
横に足す

$$E(X)=0\cdot\frac{1}{4}+1\cdot\frac{2}{4}+2\cdot\frac{1}{4}=1$$ 終

例題 2） 袋の中に赤玉 3 個と白玉 2 個が入っている。この中から 3 個の玉を同時に取り出すとき，赤玉の個数を X とする。確率変数 X の期待値を求めよ。

解答 X のとりうる値は 1，2，3 である。各値について，X がその値をとる確率を求めると

$$P(X=1)=\frac{{}_3C_1\times{}_2C_2}{{}_5C_3}=3\times1\times\frac{2\cdot1}{5\cdot4}=\frac{3}{10}$$

$$P(X=2)=\frac{{}_3C_2\times{}_2C_1}{{}_5C_3}=3\times2\times\frac{2\cdot1}{5\cdot4}=\frac{6}{10}$$

$$P(X=3)=\frac{{}_3C_3\times{}_2C_0}{{}_5C_3}=1\times1\times\frac{2\cdot1}{5\cdot4}=\frac{1}{10}$$

おさらい
${}_nC_1=n$，${}_nC_n={}_nC_0=1$
また，${}_nC_r={}_nC_{n-r}$ より
${}_5C_3={}_5C_2$，${}_3C_2={}_3C_1$

よって，X の確率分布は右の表のようになる。

X	1	2	3	計
P	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1

ゆえに，X の期待値は

$$E(X)=1\cdot\frac{3}{10}+2\cdot\frac{6}{10}+3\cdot\frac{1}{10}=\frac{18}{10}=\frac{9}{5}$$

期待値を求める手順

① 確率変数 X のとりうる値を定める

② 確率変数 X の各々の値に対応する確率を求める

③ 確率分布表を書く

④ 期待値を定義に従って計算する

□確率変数の分散と標準偏差

同じ期待値をもつ確率変数であっても，その分布は同じとは限らない。
値が期待値の近くに集中している分布もあれば，値が期待値から遠くに散らばっている分布もある。
ここでは，確率変数の値がその期待値からどの程度散らばっているかを，数量的に表すことを考えよう。
確率変数 X が右の表に示された分布に従うとする。

X	x_1	x_2	……	x_n	計
P	p_1	p_2	……	p_n	1

期待値＝平均

その期待値が m であるとすると，
このとき，X の各値と m とのへだたりの程度を表す量として

$$(x_1-m)^2,\ (x_2-m)^2,\ (x_3-m)^2,\ \cdots\cdots,\ (x_n-m)^2$$

が考えられ，$(X-m)^2$ はこれらの値をとる確率変数である。

これは
（平均との差）の２乗
つまり
（偏差）の２乗
この平均を求めれば
分散を求めることができる
（基本数Ｉと同じ）

確率変数 $(X-m)^2$ の期待値 $E((X-m)^2)$ を，
確率変数 X の **分散** といい，$V(X)$ で表す。

$V(X)$ の V は，分散を意味する英語 variance の頭文字である。

分散は、「確率分布のばらつきを示すもの」

すなわち，$V(X)$ は次の式で表される。

$$V(X)=(x_1-m)^2p_1+(x_2-m)^2p_2+\cdots\cdots+(x_n-m)^2p_n$$

また，和の記号 $\sum$ を使って表すと，次のようになる。

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

確率変数の分散

$$\begin{aligned}V(X)&=E((X-m)^2)\\&=(x_1-m)^2p_1+(x_2-m)^2p_2+\cdots\cdots+(x_n-m)^2p_n\\&=\sum_{k=1}^{n}(x_k-m)^2p_k\end{aligned}$$

偏差の２乗の期待値（平均）

分散 $V(X)$ は確率変数 $(X-m)^2$ の期待値であるから，X の測定単位が，たとえば cm であるとき，$V(X)$ の単位は cm^2 となる。そこで，X の測定単位と同じ単位である $\sqrt{V(X)}$ を散らばりの度合いを表す数値として用いることも多い。$\sqrt{V(X)}$ を X の **標準偏差** といい，$\boldsymbol{\sigma}(X)$ で表す。

確率変数の標準偏差

$$\sigma(X)=\sqrt{V(X)}$$

数Ｉのとき同様
$\sqrt{(分散)}$
$\sqrt{V(X)}$ は $V(x)$ の
正の平方根

$\sigma(X)$ の σ はギリシャ文字の小文字で
「シグマ」と読む。
標準偏差を意味する英語は
standard deviation であり，その頭文字 s
に相当するギリシャ文字が σ である。

例２） 1 個のさいころを 1 回投げたときに出る目を X とすると，X と $(X-m)^2$ の確率分布は右の表のようになる。

X	1	2	3	4	5	6	計
$(X-m)^2$	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$	計
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

したがって，X の期待値，分散，標準偏差は，次のようになる。

$$E(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{7}{2}$$

$$V(X)=\left(1-\frac{7}{2}\right)^2\cdot\frac{1}{6}+\left(2-\frac{7}{2}\right)^2\cdot\frac{1}{6}+\left(3-\frac{7}{2}\right)^2\cdot\frac{1}{6}$$
$$+\left(4-\frac{7}{2}\right)^2\cdot\frac{1}{6}+\left(5-\frac{7}{2}\right)^2\cdot\frac{1}{6}+\left(6-\frac{7}{2}\right)^2\cdot\frac{1}{6}$$
$$=\frac{1}{6}\cdot\left(\frac{25}{4}+\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4}+\frac{25}{4}\right)=\frac{1}{6}\cdot\frac{70}{4}=\frac{70}{24}=\frac{35}{12}$$

$$\sigma(X)=\sqrt{\frac{35}{12}}=\frac{\sqrt{105}}{6}$$ 終

X の分散を表す式 ① の右辺を変形してみよう。

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

$\sum_{k=1}^{n}x_kp_k=m$

$\sum_{k=1}^{n}p_k=p_1+p_2+p_3+\cdots\cdots+p_n=1$

ここで，$m=E(X)$ であるから，次の等式が成り立つ。

数学 I では
(X の分散)
$=$(X^2 の平均)$-$(X の平均)^2
２乗の平均 − 平均の２乗

分散と期待値

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

(X の分散)
$=$(X^2 の期待値)$-$(X の期待値)^2
２乗の期待値 − 期待値の２乗

別解 例 2 の $V(X)$ を，上の公式を用いて計算すると

X	1	2	3	4	5	6	計
X^2	1	4	9	16	25	36	計
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$E(X)=\frac{7}{2}$， 期待値

$E(X^2)=\frac{1}{6}\cdot(1+4+9+16+25+36)=\frac{91}{6}$ ２乗の期待値

であるから　$V(X)=E(X^2)-\{E(X)\}^2=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}$ 終

統計的な推測【確率変数の期待値と分散】 p.56〜60

青チャートB: 例題 63〜67

確率変数 X の期待値，分散，標準偏差を，それぞれ X の分布の **平均**，**分散**，**標準偏差** ともいう。標準偏差 $\sigma(X)$ は，X の分布の平均 m を中心として，X のとる値の散らばる傾向の程度を表している。**標準偏差 $\sigma(X)$ の値が小さいほど，X のとる値は，平均 m の近くに集中する**傾向にある。

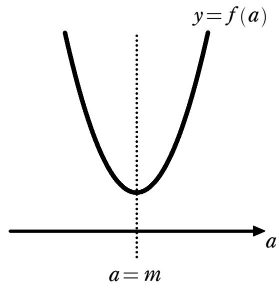
深める　上の破線のようになる理由を，分散の定義式 $V(X)=\sum_{k=1}^{n}(x_k-m)^2 p_k$ を使って説明してみよう。

標準偏差 $\sigma(X)=\sqrt{V(X)}$ の値が小さいとき，分散 $V(X)=\sum_{k=1}^{n}(x_k-m)^2 p_k$ の値が小さい。
そのとき，X の各値 x_k と平均 m について，$(x_1-m)^2$，$(x_2-m)^2$，……，$(x_n-m)^2$ の値が小さい。すなわち X のとる値が m の近くに集中している。

補足　分布の代表値の一つを a とし，分布のばらつきの関数を $f(a)$ を $f(a)=\dfrac{1}{n}\sum_{k=1}^{n}(x_k-a)^2$ と定義する。ただし，m は平均値とする。

$$\begin{aligned}
f(a)&=\frac{1}{n}\{(x_1-a)^2+(x_2-a)^2+(x_3-a)^2+\cdots\cdots+(x_n-a)^2\}\\
&=\frac{1}{n}\{na^2-2a(x_1+x_2+x_3+\cdots\cdots+x_n)+(x_1{}^2+x_2{}^2+x_3{}^2+\cdots\cdots+x_n{}^2)\}\\
&=\frac{1}{n}\{na^2-2anm+(x_1{}^2+x_2{}^2+x_3{}^2+\cdots\cdots+x_n{}^2)\}\\
&=a^2-2am+\frac{1}{n}\cdot(x_1{}^2+x_2{}^2+x_3{}^2+\cdots\cdots+x_n{}^2)\\
&=(a-m)^2+\frac{1}{n}\cdot(x_1{}^2+x_2{}^2+x_3{}^2+\cdots\cdots+x_n{}^2)-m^2
\end{aligned}$$

平方完成



よって $y=f(a)$ は，$a=m$ で最小となり，
その値は「**2乗の平均　引く　平均の2乗**」になっている。

したがって，代表値の中でも平均 m を利用した方が，
ばらつき $f(a)$ を最小で評価できる。

【まとめ】
分布のばらつきは，代表値の中でも**『平均を基準』**に置いた方が最小で評価できる。
このとき，確率変数の平均と分散の関係が導ける。
$V(X)=E(X^2)-\{E(X)\}^2$

補足　分布の代表値の一つを a とし，分布のばらつきの関数を $f(a)$ を

$f(a)=\dfrac{1}{n}\sum_{k=1}^{n}|x_k-a|$ と定義する。ただし，m は平均値とする。

$$f(a)=\frac{1}{n}\{|x_1-a|+|x_2-a|+|x_3-a|+\cdots\cdots+|x_n-a|\}$$

ここで　$x_1\leqq x_2\leqq x_3\leqq\cdots\cdots\leqq x_n$　として考える

区間 $(-\infty,\ x_1)$, $[x_2,\ x_3)$, $[x_3,\ x_4)$, ……, $[x_n,\ \infty)$ において，$f(a)$ は a の１次関数になる

区間 $(-\infty,\ x_1)$ では，

$$\begin{aligned}f(a)&=\frac{1}{n}\{-(a-x_1)-(a-x_2)-(a-x_3)-\cdots\cdots-(a-x_n)\}\\&=\frac{1}{n}\{-na+(x_1+x_2+x_3+\cdots\cdots+x_n)\}\\&=-a+\frac{1}{n}(x_1+x_2+x_3+\cdots\cdots+x_n)\\&=-a+m\end{aligned}$$

区間 $[x_1,\ x_2)$ では，

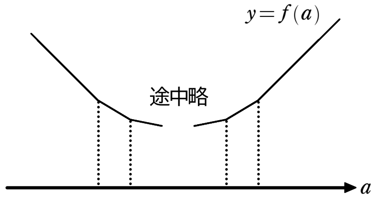
$$\begin{aligned}f(a)&=\frac{1}{n}\{(a-x_1)-(a-x_2)-(a-x_3)-\cdots\cdots-(a-x_n)\}\\&=\frac{1}{n}\{-(n-2)a+(-x_1+x_2+x_3+\cdots\cdots+x_n)\}\\&=\left(-1+\frac{2}{n}\right)a+\frac{1}{n}(-x_1+x_2+x_3+\cdots\cdots+x_n)\end{aligned}$$

区間 $[x_2,\ x_3)$ では，

$$\begin{aligned}f(a)&=\frac{1}{n}\{(a-x_1)+(a-x_2)-(a-x_3)-\cdots\cdots-(a-x_n)\}\\&=\frac{1}{n}\{-(n-4)a+(-x_1-x_2+x_3+\cdots\cdots+x_n)\}\\&=\left(-1+\frac{4}{n}\right)a+\frac{1}{n}(-x_1-x_2+x_3+\cdots\cdots+x_n)\end{aligned}$$

区間 $[x_n,\ \infty)$ では，

$$\begin{aligned}f(a)&=\frac{1}{n}\{(a-x_1)+(a-x_2)+(a-x_3)+\cdots\cdots+(a-x_n)\}\\&=\frac{1}{n}\{na-(x_1+x_2+x_3+\cdots\cdots+x_n)\}\\&=a-\frac{1}{n}(x_1+x_2+x_3+\cdots\cdots+x_n)\\&=a-m\end{aligned}$$



よって，ばらつき $y=f(a)$ は，各区間をつなぎ合わせた下に凸の折れ線となる。

$y=f(a)$ のグラフは，左側は傾き -1 の半直線で，右側は傾き $+1$ の半直線になっている

$f(a)$ の最小は，n が奇数のとき，$a=x_{\frac{n+1}{2}}$ のとき，n が偶数のとき，a が区間 $\left[x_{\frac{n}{2}},\ x_{\frac{n}{2}+1}\right]$ にあるときであり

ばらつき $f(a)$ を最小にする a は，平均値ではなく「**中央値**」である。

【まとめ】

分散は，平均を基準にした方が，コンパクトでよい。

偏差の絶対値の平均は，データの中央値が最小となり，必ずしも平均値ではない。

ちなみに　$\dfrac{1}{n}\sum_{k=1}^{n}|x_k-m|$　(mは平均) は，平均絶対偏差と呼ばれている。

（高校の統計学では，平均絶対偏差は扱いにくく，標準偏差が用いられている。）

また，ばらつき $f(a)$ を一般化して，$f(a)=\dfrac{1}{n}\sum_{k=1}^{n}|x_k-a|^p$ と定義すれば，

平均絶対偏差は $p=1$ のとき，標準偏差は $p=2$ のときである。

この一般化された $f(a)$ に対して，$\{nf(a)\}^{\frac{1}{p}}$ を L^p ノルムと言い，関数解析の分野で扱われている。