5　5回目講義の補足

5.1　連絡3件

(i) プリントはこれまでのものも含めて情報基盤センターの教育用計算機システム講義用 WWW サーバにアップしてあります（講義前日の夕方くらいにはアップしている）ので、各自必要に応じて利用して下さい。URL は http://lecture.ecc.u-tokyo.ac.jp/~chkurata/ です。

(ii) テキストの演習問題の解答や確率分布表、データはサイエンス社・新世社のＨＰからダウンロード出来ます。トップページ http://www.saiensu.co.jp/ に検索欄があるのでそこに「入門統計解析」と入力。本テキストのページが開いたら、「サポート情報あり」のアイコンをクリックする。

(iii) 練習問題集としては、村上正康・安田正實「統計学演習」培風館が手頃と思います。

5.2　確率変数

例 1. (コイン投げ)

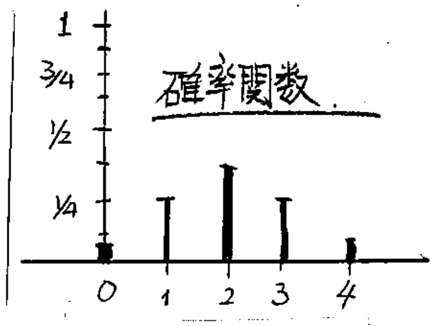
(A) コインを1回投げることを考える。結果を X で表し、

$$X = \begin{cases} 0 & (\text{裏}) \\ 1 & (\text{表}) \end{cases}$$

と定義すると、

$$\begin{cases} * \ X \text{ の取り得る値の集合} = \{0,\ 1,\} \\ * \ P(X=0) = 1/2,\ P(X=1) = 1/2 \end{cases}$$

である。



(B) コインを n 回投げることを考える。変数 X を

$$X \ = \ n \text{ 回中で表が出る回数}$$

と定義すると、前回勉強した通り

$$\begin{cases} * \ X \text{ の取り得る値の集合} = \{0,\ 1,\ \cdots,\ n\} \\ * \ P(X=x) = {}_nC_x/2^n \quad (x = 0, 1, \cdots, n) \end{cases}$$

となる。//

X= 4回投げる時の表の回数

$P(X=x) = {}_4C_x / 2^4$ $(x=0,1,2,3,4$

x	0	1	2	3	4
$P(X=x)$	1/16	4/16	6/16	4/16	1/16

定義 (上例のように、) 変数 X が次の条件

$$\begin{cases} * \ X \text{ の取り得る値の集合（以後、値域と呼ぶ）が分かっている} \\ * \ \text{各値に確率が付与されている} \end{cases}$$

を満たすとき、X を確率変数 (random variable, r.v.) と言う。データを「確率変数の実現値」と定義する。

5.3 離散型確率変数の定義

確率変数 X の値域が、次のような形の集合

$$\{x_1, x_2, \cdots\} \text{ (可算集合、これは有限集合も含む)} \tag{5.1}$$

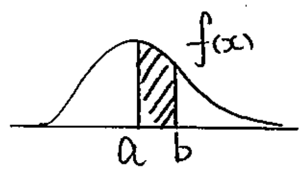
であるとき、X を離散型確率変数と言う (X は飛び飛びの値しか取らない)。X の確率変数としての性質は、

$$P(X = x_k) = f(x_k) \quad (k = 1, 2, \cdots) \tag{5.2}$$

によって定まる。これを X の確率分布と言う。

5.4 連続型確率変数の定義

$P(a \leq X \leq b) =$



確率変数 X の値域が、

$$(-\infty, \infty)、[0, \infty) \text{ などの区間} \tag{5.3}$$

であり、かつある非負の関数 $f(x)$ が存在して

$$P(a \leq X \leq b) = \int_a^b f(x)dx \quad \text{for all } a \leq b \tag{5.4}$$

が成立するとき、X を連続型確率変数と言う。X の確率変数としての性質は、関数 $f(x)$ によって定まる。$f(x)$ を確率密度関数 (probability density function, pdf) と言う。

5.5 コイン投げ

コインを n 回投げることを考える。変数 X を

$$X = n \text{ 回中で表が出る回数}$$

と定義すると、X は

$$P(X = x) = f(x) = \frac{{}_nC_x}{2^n} \quad (x = 0, 1, \cdots, n) \tag{5.5}$$

なる確率分布に従う離散型確率変数である。

定理 5.1. $E(X) = n/2$、$V(X) = n/2^2$ である。

証明はより一般的な形で次回行う。

5.6 期待値と分散の練習

例 2. (宝くじ) 次表のような宝くじを考える。発行枚数は 10000 枚とする。

当選金 (円)	0	100	1000	10000
枚数	9000	800	150	50

この宝くじを 1 枚購入したときの当選金を X とおけば、抽選前の X は次の分布を持つ離散型確率変数とみなせる。

x_k	0	100	1000	10000
$P(X=x_k)$	0.9	0.08	0.015	0.005

X の期待値 $\mathbf{E}(X)$ は定義により、

$$\mathbf{E}(X) = 0\times 0.9 + 100\times 0.08 + 1000\times 0.015 + 10000\times 0.005 = 73(\text{円}) \tag{5.6}$$

となる。

期待値の定義に基づいて、$\mathbf{E}[0.8X+2]$ を求める。$g(X)=0.8X+2$ とおくと、定義より

$$\begin{aligned}
\mathbf{E}[0.8X+2] &= \mathbf{E}[g(X)] && (5.7)\\
&= g(0)\times P(X=0) + g(100)\times P(X=100) + g(1000)\times P(X=1000) + g(10000)\times P(X=10000) &&\\
&= (0.8\times 0+2)\times 0.9 + (0.8\times 100+2)\times 0.08 &&\\
&+ (0.8\times 1000+2)\times 0.015 + (0.8\times 10000+2)\times 0.005 && (5.8)\\
&= 60.4 && (5.9)
\end{aligned}$$

となる.

期待値の線形性を用いることにより、期待値 $\mathbf{E}[0.8X]$ の計算はずっと簡単であり、

$$\mathbf{E}[0.8X+2] = 0.8\mathbf{E}(X) + 2 = 0.8\times 73 + 2 = 60.4. \tag{5.10}$$

と求めればよい。

分散を計算する。$\mathbf{E}(X)=73$ であったから、定義により、

$$\begin{aligned}
\mathbf{V}(X) &= (0-73)^2\times 0.9 + (100-73)^2\times 0.08 + (1000-73)^2\times 0.015\\
&\quad + (10000-73)^2\times 0.005 = 510471.0 \qquad (5.11)
\end{aligned}$$

となる。$\mathbf{D}(X)=\sqrt{510471.0}=714.5$ である。□

例 3. (分散の計算練習) 2 つの確率変数 X, Y がそれぞれ次のような確率分布に従うとする.

	k	0	1	2	3	4	5	6
散らばり大	$P(X=k)$	0.1	0.1	0.2	0.2	0.2	0.1	0.1
〃 〃 小	$P(Y=k)$	0	0.1	0.2	0.4	0.2	0.1	0

$$\mathbf{E}(X) = 0\times 0.1 + 1\times 0.1 + 2\times 0.2 + 3\times 0.2 + 4\times 0.2 + 5\times 0.1 + 6\times 0.1 = 3, \tag{5.12}$$

$$\mathbf{E}(X^2) = 0^2\times 0.1 + 1^2\times 0.1 + 2^2\times 0.2 + 3^2\times 0.2 + 4^2\times 0.2 + 5^2\times 0.1 + 6^2\times 0.1 = 12, \tag{5.13}$$

$$\mathbf{V}(X) = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2 = 12 - 3^2 = 3. \tag{5.14}$$

同様に $\mathbf{V}(Y)=1.2$ である (各自で)。$\mathbf{V}(X) > \mathbf{V}(Y)$ が成り立っている。 □

5.7 Chebyshev の不等式と応用例

[Chebyshev の不等式] 確率変数 X の平均を μ、分散を σ^2 で表す：

$$E(X) = \mu, \quad V(X) = \sigma^2.$$

このとき、任意の $k \geq 0$ に対して次の不等式が成り立つ。

$$P\Big(|X-\mu| \geq k\sigma\Big) \leq \frac{1}{k^2}, \tag{5.10}$$

$$P\Big(|X-\mu| \leq k\sigma\Big) \geq 1 - \frac{1}{k^2} \tag{5.11}$$

(ここで、

$$P\Big(|X-\mu| \leq k\sigma\Big) = P\Big(\mu - k\sigma \leq X \leq \mu + k\sigma\Big) = P\Big(-k \leq \frac{X-\mu}{\sigma} \leq k\Big)$$

なることに注意する。)

特に、$k = 2, 3$ として、

$$P\Big(\mu - 2\sigma \leq X \leq \mu + 2\sigma\Big) \geq \frac{3}{4} = 0.75 \tag{5.12}$$

$$P\Big(\mu - 3\sigma \leq X \leq \mu + 3\sigma\Big) \geq \frac{8}{9} \approx 0.889 \tag{5.13}$$

が成立する。

例 4. (コイン投げ) 表が出る確率が 0.5 であるようなコインを 10000 回投げる。表の出る回数を X とおけば、X の確率分布は

$$P(X = x) = f(x) = \frac{{}_{10000}C_x}{2^{10000}} \quad (x = 0, 1, \cdots, 10000)$$

であり、

$$\mu = E(X) = 10000/2 = 5000, \ \sigma^2 = V(X) = 10000/4 = 2500, \ \sigma = 50$$

である。

Chebyshev の不等式を用いて、$P(4800 \leq X \leq 5200)$ を求めてみよう。

$$\begin{aligned} P\Big(4800 \leq X \leq 5200\Big) &= P\Big(5000 - 4 \times 50 \leq X \leq 5000 + 4 \times 50\Big) \\ &= P\Big(\mu - 4\sigma \leq X \leq \mu + 4\sigma\Big) \\ &\geq 1 - \frac{1}{4^2} = \frac{15}{16} \approx 0.9375 \end{aligned} \tag{5.14}$$

実際にはこの確率は 0.99994 であるから (後日)、Chebyshev の不等式はかなり粗いものであることが分かる。しかし、この不等式はどのような確率分布に対しても成立するという長所がある。//