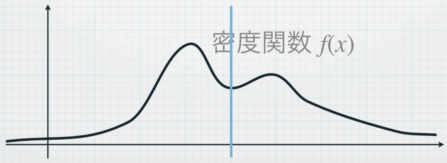
確 率 論 第７回
-- 期待値と分散 --

情報工学科 山本修身

確率分布の特徴をとらえる

* 確率分布が与えられたとき，その特徴を抽出することによって，その確率分布をおおまかに理解することができる．
* 特徴として一番単純なのは平均（期待値）である．



期待値とは（連続的確率変数の場合）

* **期待値** (expectation) は，**平均** (mean) とも呼ばれる．確率変数 X の密度関数を$f(x)$としたとき，

$$E(X) = \int_{-\infty}^{\infty} f(x)xdx$$

と表現される．

期待値とは（離散的確率変数の場合）

* 離散的な確率変数の場合には，積分では表現できない．確率関数を$p(x)$とおけば，

$$E(X) = \sum_{i=1}^{\infty} p(x_i)x_i$$

と表現できる．要は，それぞれの値とそれに対する出現確率を掛け合わせてすべて足せば良い．

期待値の計算例 (1)

* サイコロを振ったとき，出た目の数に対応する確率変数をXとする．この確率変数の平均は，

$$E(X) = 1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}$$
$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

と計算することができる．

期待値の計算例 (2)

* 0と2の間の実数をランダムに出力する確率変数をXとする．どの数も同じように確からしく出現するとする（一様分布）．密度関数は
f(x) = 1/2 (0 ≤ x ≤ 2)
と表現される．この確率変数の期待値は，つぎのようになる．

$$E(X) = \int_0^2 \frac{1}{2}xdx = \left[\frac{x^2}{4}\right]_0^2 = \frac{4-0}{4} = 1.$$

期待値の計算例 (3)

* 宝くじの期待値．以下のような宝くじを1枚買うことによって得られる金の期待値は以下のとおり．

$$E(X) = 6000 \times 10^4 \times \frac{5}{1000 \times 10^4} + 1500 \times 10^4 \frac{10}{1000 \times 10^4} + \cdots = 139.58$$

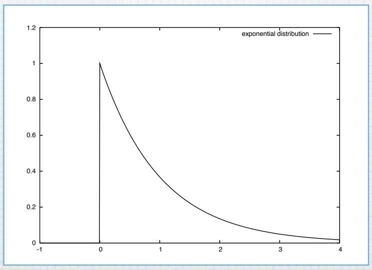
全部で1000万本．これ以外ははずれ

6000万	5	1500万	10	1000万	10
100万	60	10万	595	7万	90
1万	2000	3000	10万	400	100万

確率分布の例：指数分布 (再掲)

* 指数分布はつぎのような密度関数を持つ確率分布である：

$$f(x) = \lambda \mathrm{e}^{-\lambda x} \quad (x > 0)$$



(再掲) 指数分布の性質 $\int_0^y \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^y = 1 - e^{-\lambda y}$

* 密度関数を積分することにより，分布関数は，

$$F(y) = 1 - \mathrm{e}^{-\lambda y} \quad (y > 0)$$

である．また，確率変数Xがx以上であるという条件のもとで，[x, x + Δx]に入る確率は

$$\begin{aligned} P\,(x < X \le x + \Delta x \mid x < X) &= \frac{P(x < X \le x + \Delta x)}{P(x < X)} \\ &= \frac{F(x + \Delta x) - F(x)}{1 - F(x)} \\ &= \frac{f(x)\Delta x}{1/\lambda f(x)} = \lambda \Delta x \end{aligned}$$

と計算できる．

指数分布従う現象

* 色々な現象があるが，有名なのは，放射性の原子が崩壊するまでの時間の分布である．時間とともに崩壊していない確率は小さくなっていく．ちょうど崩壊していない確率が1/2になる時間を半減期と呼ぶ．これはこの分布の性質から ln2 / λ となる．

$$\begin{aligned} 1 - e^{-\lambda x} &= 1/2 \\ e^{-\lambda x} &= 1/2 \\ -\lambda x &= -\ln 2 \\ x &= \ln 2/\lambda \end{aligned}$$

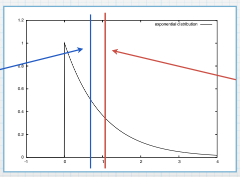
期待値の計算例 (4)

* 指数分布に従う確率変数の期待値．

$$E(X) = \int_0^\infty \lambda \mathrm{e}^{-\lambda x} x dx \qquad f(x) = \lambda \mathrm{e}^{-\lambda x} \quad (x > 0)$$

$$E(X) = [-\mathrm{e}^{-\lambda x} \cdot x]_0^\infty + \int_0^\infty \mathrm{e}^{-\lambda x} dx = \int_0^\infty \mathrm{e}^{-\lambda x} dx = \left[-\frac{e^{-\lambda x}}{\lambda} \right]_0^\infty = \frac{1}{\lambda}$$

$x = \ln 2/\lambda$
半減期



平均

λ = 1の場合

幾何分布（再掲）

* 幾何分布は離散分布でありながら，とる値が無限個存在する分布である．
* 1枚のコインを何回も振る試行を考える（これをベルヌーイ試行と呼ぶ）．このとき，表が最初にでるまでの振った回数の分布である．確率関数はつぎのようになる．

H: p

T: $q = 1 - p$

$$P(X = k) = q^{k-1} p \quad (k = 1, 2, 3, \ldots)$$

ただし，$q = 1 - p$

∞回振る

1	2	3	4	5	6	7	8	…
T	T	H	H	T	H	T	H	…

幾何分布の性質（再掲）

- 確率変数は無限種類の値をとるが，確率を全部足すと1になる．

$$\sum_{i=1}^{\infty} P(X=k) = \sum_{i=1}^{\infty} q^{k-1}p = \frac{p}{1-q} = 1$$

$$q = 1-p$$

期待値の計算例 (5)

- 幾何分布の平均値
無限級数の和を計算する必要がある．

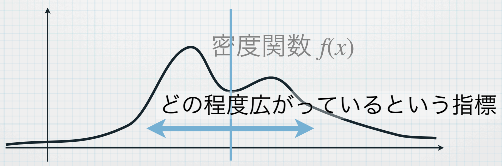
$$p(k) = pq^{k-1} \quad (k=1,2,\cdots)$$

$$\begin{aligned} E(X) &= \sum_{i=1}^{\infty} pq^{k-1}k = \sum_{i=1}^{\infty} p\left.\frac{dx^k}{dx}\right|_{x=q} = p\frac{d}{dx}\left.\left(\sum_{i=1}^{\infty} x^k\right)\right|_{x=q} \\ &= p\frac{d}{dx}\left.\left(\frac{x}{1-x}\right)\right|_{x=q} = p\frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

表の出る確率が小さければ表の出るまでの回数は多くなる．

分散とは

- 平均は確率変数の分布を端的に示す指標であるが，それだけでは，どのくらいの広がりをもって分布しているのかがわからない．
- そこで，平均の周りにどの程度分布するかの指標として分散 (variance) がある．平均と合わせて，分布の良い指標となる．



分散の定義

- **分散 (variance)** は平均（期待値）μ からそれぞれの値がどのくらい離れているかを測る指標である．連続的確率変数の場合，密度関数f(x)を用いて，

$$V(X) = \int_{-\infty}^{\infty} f(x)(x-\mu)^2 dx$$

と書くことができる．離散的確率変数の場合，

$$V(X) = \sum_{i=1}^{\infty} p(x_i)(x_i - \mu)^2$$

となる．

分散の計算法

- 分散をそのまま計算しようとすると，かなりの計算量となる場合が多い．一般の分散について以下の性質が成り立つので，これを利用することにより，多少計算が楽になる．

$$V(X) = \int_{-\infty}^{\infty} f(x)x^2 dx - \mu^2$$

分散に関する公式の証明

- 前述の性質はつぎのように証明できる．

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} f(x)(x-\mu)^2 dx = \int_{-\infty}^{\infty} f(x)(x^2 - 2\mu x + \mu^2) dx \\ &= \int_{-\infty}^{\infty} f(x)x^2 dx - 2\mu \underbrace{\int_{-\infty}^{\infty} f(x)x dx}_{=\mu} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x) dx}_{=1} \end{aligned}$$

$$V(X) = \int_{-\infty}^{\infty} f(x)x^2 dx - \mu^2$$

分散に関する公式

- 前述の性質は次の式のように表現される.

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

「分散とは二乗の平均から平均の二乗をひいたものである．」

標準偏差とは

- **標準偏差 (standard deviation)** とは分散の平方根のことである．すなわち,

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

である．分散は次元が元の確率変数と異なるので，そのまま足したり引いたりすることはできない．それに対して標準偏差は確率変数と同じ次元となり，扱いやすくなる.

分散の計算例 (1)

- サイコロの例：分散を計算するためにまず二乗の平均を計算する：

$$\begin{aligned} E(X^2) &= 1^2\cdot\frac{1}{6} + 2^2\cdot\frac{1}{6} + 3^2\cdot\frac{1}{6} + 4^2\cdot\frac{1}{6} + 5^2\cdot\frac{1}{6} + 6^2\cdot\frac{1}{6} \\ &= \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6} \end{aligned}$$

これより，分散はつぎのようになる.

$$V(X) = E(X^2) - E(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = 2.92$$

さらに標準偏差はつぎのようになる. $\sqrt{35/12} = 1.71$

分散の計算例 (2)

- 0と2の間のランダムな数を一様に発生させる確率変数 X の分散の計算.
- この場合，連続的な分布なので，積分になる．まず二乗の平均を求めればつぎのようになる.

$$E(X^2) = \int_0^2 x^2\frac{dx}{2} = \left[\frac{x^3}{6}\right]_0^2 = \frac{2^3}{6} = \frac{4}{3}$$

これより，分散はつぎのようになる.

$$V(X) = \frac{4}{3} - 1^2 = \frac{1}{3} \quad \therefore \sigma = \sqrt{3}/3 = 0.577$$

分散の計算例 (3)

- 幾何分布の分散の計算．まず，二乗の平均は,

$$\begin{aligned} E(X^2) &= \sum_{i=1}^{\infty} pq^{k-1}k^2 = \sum_{i=1}^{\infty} p\frac{d^2}{dx^2}x^{k+1}\Big|_{x=q} - E(X) \\ &= p\frac{d^2}{dx^2}\left(\sum_{i=1}^{\infty} x^{k+1}\right)\Bigg|_{x=q} - \frac{1}{p} \\ &= p\frac{d^2}{dx^2}\left(\frac{x^2}{1-x}\right)\Bigg|_{x=q} - \frac{1}{p} = p\frac{2}{(1-q)^3} - \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p} \end{aligned}$$

これより,

$$V(X) = \frac{2}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

チェビシェフの不等式

- チェビシェフの不等式は，平均＋σ 標準偏差という限界よりも上または平均ーσ 標準偏差よりも下に出てくる確率の上界を見積もるための不等式である．かなりのどんぶり勘定であるが，分布を仮定する必要がないので便利である.

$$\frac{1}{a^2} \geq P(|X-\mu| \geq a\sigma)$$

ただし，aは任意の正数である.

チェビシェフの不等式の証明 (1)

- まず，つぎの式が定義より成り立つ.

$$\sigma^2 = \int_{\infty}^{\mu-a\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-a\sigma}^{\mu+a\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

これより，

$$\sigma^2 \geq \int_{\infty}^{\mu-a\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

さらに $X \leq \mu - a\sigma, X \geq \mu + a\sigma$ という条件下では，以下の式が成立する：

$$(X-\mu)^2 \geq a^2\sigma^2$$

チェビシェフの不等式の証明 (2)

- 2つの式,

$$\sigma^2 \geq \int_{\infty}^{\mu-a\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

$$(X-\mu)^2 \geq a^2\sigma^2$$

より，

$$\begin{aligned}\sigma^2 &\geq \int_{\infty}^{\mu-a\sigma} a^2\sigma^2 f(x)dx + \int_{\mu+a\sigma}^{\infty} a^2\sigma^2 f(x)dx \\ &= a^2\sigma^2 \{P(X \leq \mu - a\sigma) + P(X \geq \mu + a\sigma)\}\end{aligned}$$

これより，

$$\frac{1}{a^2} \geq P(|X-\mu| \geq a\sigma)$$

チェビシェフの不等式の応用例

例　0 から 2 の間から一様にランダムに実数を出力する確率変数 X について考えると，平均は $\mu = 1$ であり，標準偏差は $\sqrt{3}/3$ であるから，

$$\mu + 2\sigma = 1 + 2\frac{\sqrt{3}}{3} > 2, \quad \mu - 2\sigma = 1 - 2\frac{\sqrt{3}}{3} < 0, \tag{29}$$

となり，この範囲には出現する確率は 0 となる．チェビシェフの不等式では 1/4 以下であるが，それよりも小さくなっており，チェビシェフの不等式は成り立っている．

$$\frac{1}{4} \geq P(|X-\mu| \geq 2\sigma)$$