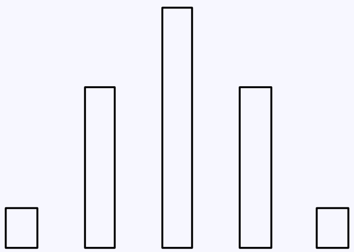
1.2.1 確率密度

4月18日（金）

鈴木直也

確率密度（1）

- 連続変数について確率を考えたいときに、今までのような離散的な事象集合に対して定義される確率分布を用いることはできず、測度の概念[1] が必要である。



注1：測度（そくど）とは与えられた集合の部分集合に対して "大きさ"、"容積"、"確率" などといった数を割り当てる関数である。

確率密度(2)

確率変数xが連続確率変数のとき、
特定された点xでの確率 を与えることができない。
そこで、x の周囲を含めた区間において、
確率変数xの与える値を考えることにする。
x が区間$(x, x+\delta x)$ 、$(\delta x \to 0)$にある確率が
$p(x)\delta x$ で与えられるとき$p(x)$ を確率密度と呼ぶ。

確率密度(3)

- 確率密度$p(x)$において確率が意味を持つのは区間で考えたとき、つまり$p(x)$における区間(a,b)の面積である。
  この時xが区間(a,b)にある確率は
  $$p(x\in(a,b))=\int_a^b p(x)dx$$
  で与えられる。

確率密度$p(x)$ の条件

- 確率は非負でxは実数値を取らなければならないから確率密度 $p(x)$ は以下の条件を満たす必要がある。

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

確率密度(4)

- 関数$f(x)$に対して変数変換$x=g(y)$を考えたとき

  $f(x)$ は$\tilde{f}(y)=f(g(y))$ となる。

  確率密度 $p_x(x)$ に対応する変換後の変数yに関する

  密度$p_y(y)$ を考える。

  区間$(x, x+\delta x)$に入る観測値はδxが小さければ

  区間$(y, y+\delta y)$に入り、$p_x(x)\delta x \simeq p_y(y)\delta y$となるので

$$p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|$$

$$= p_x(g(y))\left|g'(y)\right|$$

  となる。

累積分布関数

- 連続変数 x について確率を考えるとき区間で考える。
  これを一般化したものが累積分布関数である。
  累積分布関数では確率変数x
  が区間 $(-\infty, z)$に入る確率が

$$P(z) = \int_{-\infty}^{z} p(x)dx$$

  で定義される。

多変数確率密度

- いくつかの連続変数$x_1, \cdots\cdots, x_D$があるとき, これらをまとめてベクトルxであらわすと、同時分布 $p(\mathrm{x}) = p(x_{1,} \cdots, x_D)$
  を定義することができる。このときxがxを含む無限小の体積要素δxにはいる確率は$p(\mathrm{x})\delta \mathrm{x}$で与えられる。
  $p(\mathrm{x})$ も確率であるので

$$p(\mathrm{x}) \geq 0$$

$$\int p(\mathrm{x}) d\mathrm{x} = 1$$

  を満たす必要がある。これは離散型と連続型の変数が組み合わさっている場合も同時確率分布を考えることができる。

確率の加法･乗法定理およびベイズの定理

- 確率密度や離散変数や連続変数の組み合わせに対しても適用可能である。

  x, yを実変数として

$$p(x) = \int p(x, y)dy$$

$$p(x, y) = p(y \mid x)p(x)$$

  をとる。

確率の加法定理：2 つの事象 A と B について，一般に，
P{A ∪ B} = P{A} + P{B} − P{A ∩ B}
確率の乗法定理：事象 A が起こったという条件のもとで
事象 B の起こる確率を，A のもとでの
B の 条件付き確率 といい P{B | A}で表す。
このときPr{A ∩ B} = Pr{A} ・ Pr{B | A}
ベイズの定理：P(A) > 0 ならば、
P(B|A)=P(A|B) ＊ P(B) / P(A)

期待値と分散

期待値

- ある関数 $f(x)$ の確率分布の下での平均値を $f(x)$ の期待値と呼び、$E[f]$ で表す。

  離散変数の場合は確率質量関数で重みをつけた総和 $E[f]=\sum_{x} p(x)f(x)$ で与えられる。

  連続変数の場合は確率密度関数で重みをつけた積分 $E[f]=\int p(x)f(x)dx$ で与えられる。

期待値の近似

- 離散変数、連続変数どちらの場合も確率分布や確率密度から得られたN（有限）個の測定値を用いて近似することができる。

$$E[f] \simeq \frac{1}{N}\sum_{n=1}^{N} f(x_n)$$

さいころの出目	出た回数
1	179
2	165
3	173
4	168
5	158
6	157
期待値	3.432

生徒	身長
生徒1	170.7
生徒2	157.3
生徒3	173.5
生徒4	165.2
生徒5	183.9
:	
生徒N-1	191.3
生徒N	148.7
期待値	171.487

多変数関数の期待値

- 多変数関数の期待値の場合、どの変数に対して期待値をとるかは

$$E_x[f(x,y)]$$

のように添え字を用いて表す。このとき$E_x[f(x,y)]$は y の関数になる。

条件付き期待値

- 条件付確率についても期待値を定義することができそのときの期待値を条件付き期待値と呼ぶ。

離散変数の場合

$$E_x[f \mid y] = \sum_x p(x \mid y) f(x)$$

で与えられる。

連続変数の場合も同様に

$$E_x[f \mid y] = \int p(x \mid y) f(x) dx$$

で与えられる。

分散

ある関数$f(x)$の分散は測定値とその平均の差を２乗したものの平均で定義される。

$$\mathrm{var}[f] = E[(f(x) - E[f(x)])^2]$$

また、２乗を展開して整理すると

$$\mathrm{var}[f] = E[f(x)^2] - E[f(x)]^2$$

となる。

離散変数の共分散

2 組の対応するデータ間での、
偏差の積の平均値として定義される。
離散変数の場合

$$\begin{aligned}\mathrm{cov}[x, y] &= E_{x,y}[\{x - E[x]\}\{y - E[y]\}] \\ &= E_{x,y}[xy] - E[x]E[y]\end{aligned}$$

で与えられる。

連続変数の共分散(1)

いくつかの連続変数をまとめてベクトルx、yとしたとき共分散は行列

$$\mathrm{cov}[\mathrm{x},\ \mathrm{y}] = E_{\mathrm{x},\mathrm{y}}[\{\mathrm{x} - E[\mathrm{x}]\}\{\mathrm{y}^{\mathrm{T}} - E[\mathrm{y}^{\mathrm{T}}]\}]$$

$$= E_{\mathrm{x},\mathrm{y}}[\mathrm{x}\ \mathrm{y}] - E[\mathrm{x}^{\mathrm{T}}]E[\mathrm{y}^{\mathrm{T}}]$$

で与えられる。

連続変数の共分散(2)

- また、ベクトルxの成分間の共分散を表すのには

$$\mathrm{cov}[\mathrm{x}] \equiv \mathrm{cov}[\mathrm{x}, \mathrm{x}]$$

とかく