No.11 連続型確率分布

NAKADA Masayuki

Kobe University Secondary School

October 8, 2021

離散と連続

- **離散量**:
  さいころの目，年齢，所得のように，有限個または無限個でもとびとびの値をとる量のこと．離散量をとる確率変数を，**離散型確率変数**といい，その分布を**離散分布**という．
- **連続量**:
  身長，体重のように，連続的な値をとる量のこと．連続量をとる確率変数を，**連続型確率変数**といい，その分布を**連続分布**という．

実数に値をとる量は連続量である．
母集団が連続分布に従うものであっても，標本は無限にとることができない．
得られた標本の分布の様子をヒストグラムによって整理することがあるが，これは後述する, 確率密度関数の分布曲線の標本による近似とみなすことができる．

連続型確率分布

e.g. ボール投げの飛距離を確率変数 X[m] とする．X は（非負の）実数値をとり得る確率変数であるから，連続型確率変数である．
ボール投げにおいて，記録が 20.0[m] である確率は，$P(X = 20.0)$ ではない．記録が 20.0[m] であるとは，小数第 2 位を四捨五入した結果であろうから，幅をもった範囲 $19.95 \leqq X < 20.05$ として扱うべきである．したがって，記録が 20.0[m] である確率は，$P(19.95 \leqq X < 20.05)$ である．
なお，$X = 20.0$ とは，「X がちょうど 20.0 に等しい」という意味である．しかし，ほとんどの場合，X は 20.0 ちょうどにはならず，ややずれた値であり，厳密な意味で $X = 20.0$ となることは滅多にない．よって，$P(X = 20.0) =$ 　である．

連続型確率変数においては，ある特定の値をとる確率は，常に 0 である．そのため，連続型確率変数では，変数の値がある範囲に入る確率を考える．

連続型確率分布

e.g. ボール投げの飛距離を確率変数 X[m] とする．X は（非負の）実数値をとり得る確率変数であるから，連続型確率変数である．
ボール投げにおいて，記録が 20.0[m] である確率は，$P(X = 20.0)$ ではない．記録が 20.0[m] であるとは，小数第 2 位を四捨五入した結果であろうから，幅をもった範囲 $19.95 \leqq X < 20.05$ として扱うべきである．したがって，記録が 20.0[m] である確率は，$P(19.95 \leqq X < 20.05)$ である．
なお，$X = 20.0$ とは，「X がちょうど 20.0 に等しい」という意味である．しかし，ほとんどの場合，X は 20.0 ちょうどにはならず，ややずれた値であり，厳密な意味で $X = 20.0$ となることは滅多にない．よって，$P(X = 20.0) = 0$ である．

連続型確率変数においては，ある特定の値をとる確率は，常に 0 である．そのため，連続型確率変数では，変数の値がある範囲に入る確率を考える．

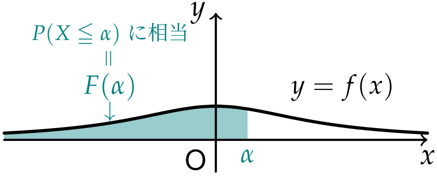
確率密度関数と確率分布関数

確率密度関数と確率分布関数

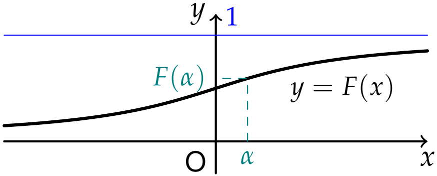
- 以下の性質をもつ関数 $f(x)$ を**確率密度関数**という．
  1. $f(x) \geqq 0$（確率は負の値をとらない）
  2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$（全事象の確率は 1）
- 確率密度関数 $f(x)$ に対し，次で定義される関数 $F(x)$ を（累積）**分布関数**という:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

※ やや不正確な言い方になるが，確率密度関数は，相対度数によるヒストグラム，（累積）分布関数は累積相対度数によるヒストグラムに対応するイメージである．



確率密度関数 $y = f(x)$



（累積）分布関数 $y = F(x)$

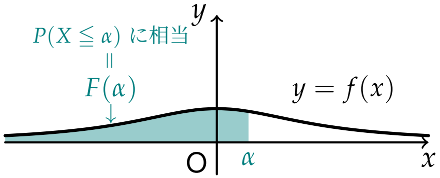
確率密度関数と確率分布関数

確率密度関数と確率分布関数

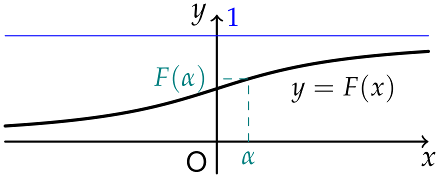
- 以下の性質をもつ関数 $f(x)$ を**確率密度関数**という．（定義域を I とすると）
  1. $f(x) \geqq 0$（確率は負の値をとらない）
  2. $\displaystyle\int_I f(x)\,dx = 1$（全事象の確率は 1）
- 確率密度関数 $f(x)$ に対し，次で定義される関数 $F(x)$ を（累積）**分布関数**という:

  $F(x) = \displaystyle\int_{I^x} f(t)\,dt$（$I^x = I \cap \{x$ 以下$\}$）

※ やや不正確な言い方になるが，確率密度関数は，相対度数によるヒストグラム，（累積）分布関数は累積相対度数によるヒストグラムに対応するイメージである．



確率密度関数 $y = f(x)$



（累積）分布関数 $y = F(x)$

連続型確率分布と平均，分散

区間 I に値をとる確率変数 X の確率密度関数を $f(x)$ とするとき，X の平均 $\mu = E(X)$ と分散 $\sigma^2 = V(X)$ は，次の式で定義される．

$$\boxed{E(X) = \int_I x f(x)\,dx, \qquad V(X) = \int_I (x-\mu)^2 f(x)\,dx}$$

また，標準偏差 $\sigma(X) = \sqrt{V(X)}$ で定める．

連続型と離散型の式の比較

連続型	離散型
確率密度関数　$f(x)$	確率関数　$f(x_i)$
$P(a \leqq X \leqq b) = \int_a^b f(x)\,dx$	$P(a \leqq X \leqq b) = \sum_{a \leqq x_i \leqq b} f(x_i)$
$E(X) = \int_I xf(x)\,dx$	$E(X) = \sum_i x_i f(x_i)$
$V(X) = \int_I (x-\mu)^2 f(x)\,dx$	$V(X) = \sum_i (x_i-\mu)^2 f(x_i)$

主な確率分布

一様分布 $U(\alpha,\beta)$

確率変数 X が区間 $[\alpha,\beta]$ 上の値を等確率でとる分布である．

確率密度関数 $f(x)=\dfrac{1}{\beta-\alpha}$ （$\alpha \leqq x \leqq \beta$），0（その他の x）

一様分布については，確率密度関数の定義域を最初から $[\alpha,\beta]$ に制限して考えることも多い（教科書はこっちの流儀）が，ここではこだわらない．

平均と分散 $E(X)=\dfrac{\alpha+\beta}{2}$，$V(X)=\dfrac{(\beta-\alpha)^2}{12}$

正規分布 $N(\mu,\sigma)$

正規分布はさまざまな現象に自然に現れる分布であるのみならず，十分大きいデータサイズの標本の標本平均の分布が正規分布に近づいていく（**中心極限定理**）など，統計学における親玉的分布である．別名ガウス分布．

確率密度関数 $\boxed{f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)}$

平均と分散 $\boxed{E(X)=\mu,\ \ V(X)=\sigma^2}$

ここで，$e = 2.7182\ldots$ はネイピア数と呼ばれる無理数で，極限 $e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$ で与えられる．

特に，$\mu = 0$，$\sigma = 1$ である正規分布 $N(0,1)$ を**標準正規分布**と呼ぶ．他の正規分布は，確率変数の標準化 $Z = \dfrac{X-\mu}{\sigma}$ によって，標準正規分布に帰着させて計算する．

正規分布の性質と取扱いについては，後ほど詳しく取り扱う．

ガンマ分布 $G_A(\alpha, \beta)$

標本の分散の分布（χ^2 分布）は，ガンマ分布の一種である．

確率密度関数 $f(x) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$ $(0 < x < \infty)$

平均と分散 $E(X) = \alpha\beta$, $V(X) = \alpha\beta^2$

$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x}\, dx$（$s > 0$）で与えられる関数を**ガンマ関数**という．

ガンマ関数の性質（参考）

$$\Gamma(1) = 1, \ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \ \Gamma(s) = (s-1)\Gamma(s-1)$$

特に，$\Gamma(n) = (n-1)!$（$n \in \mathbb{N}$）により，ガンマ関数を階乗関数 $n!$ の連続化として理解することができる．

ベータ分布 $B_E(\alpha, \beta)$

ベイズ統計（観測によって分布を更新する推測手法）を扱いだすとお世話になる.

確率密度関数 $f(x) = \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $(0 < x < 1)$

平均と分散 $E(X) = \dfrac{\alpha}{\alpha+\beta}$, $V(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

2 変数関数 $B(\alpha,\beta) = \displaystyle\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$ $(\alpha,\beta > 0)$ を**ベータ関数**という.

ベータ関数の性質（参考）

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$\Gamma(n) = (n-1)!$ $(n \in \mathbb{N})$ だったから, ベータ関数は二項係数の逆数の連続化と思える: $B(m,n) = \dfrac{1}{{}_{m+n-2}\mathrm{C}_{m-1}}$ $(m, n \in \mathbb{N})$.

標本調査の考え方

標本平均の平均と分散について