教養科目／B 自然の摂理の探求

2026 年度 統計学（集中１期）

第８回

前回までの内容（2項分布）

・**2 項分布**

X の確率関数 $P(x)$ が

$$P(x) = {}_n\mathrm{C}_x\, p^x (1-p)^{n-x} \qquad (x = 0, 1, \cdots, n)$$

となるような確率分布を**２項分布**とよぶ.

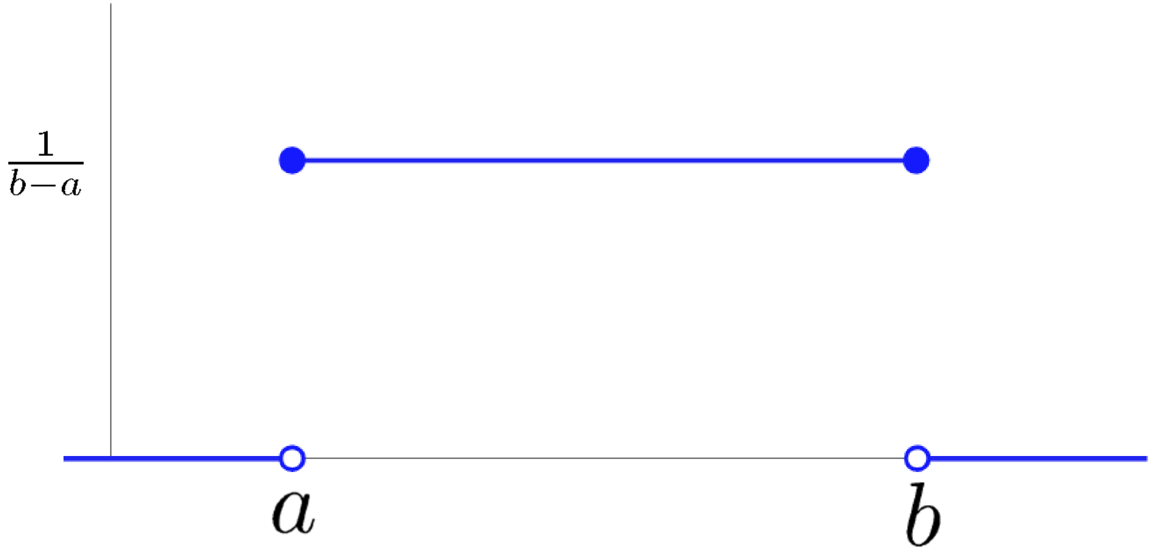
・また，X は**２項分布 $\mathrm{B}(n,p)$ に従う**といい，$X \sim \mathrm{B}(n,p)$ と表す.

前回までの内容（一様分布）

連続型確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b), \\ 0 & (\text{その他}) \end{cases}$$

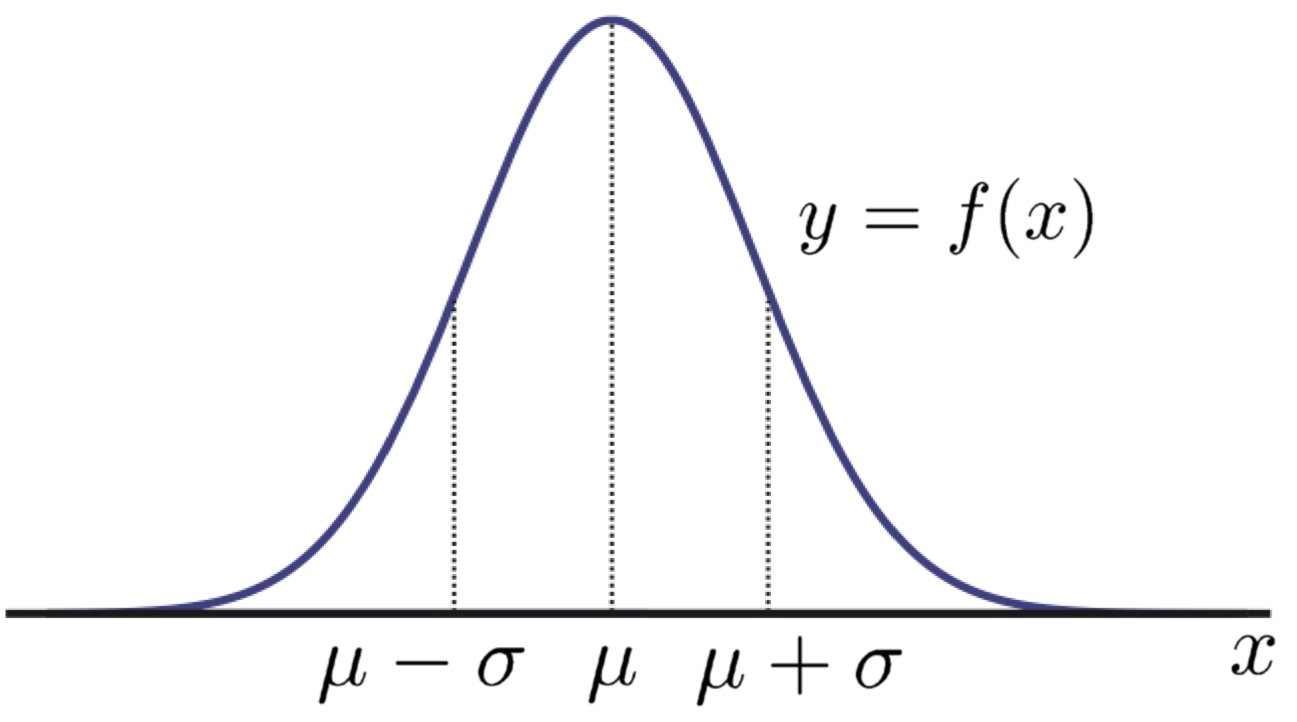
のとき，X の確率分布を $[a,b]$ 上の**一様分布**といい，X が一様分布に従うことを記号で $X \sim \mathrm{U}(a,b)$ と表す．



前回までの内容（正規分布）

正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)$$



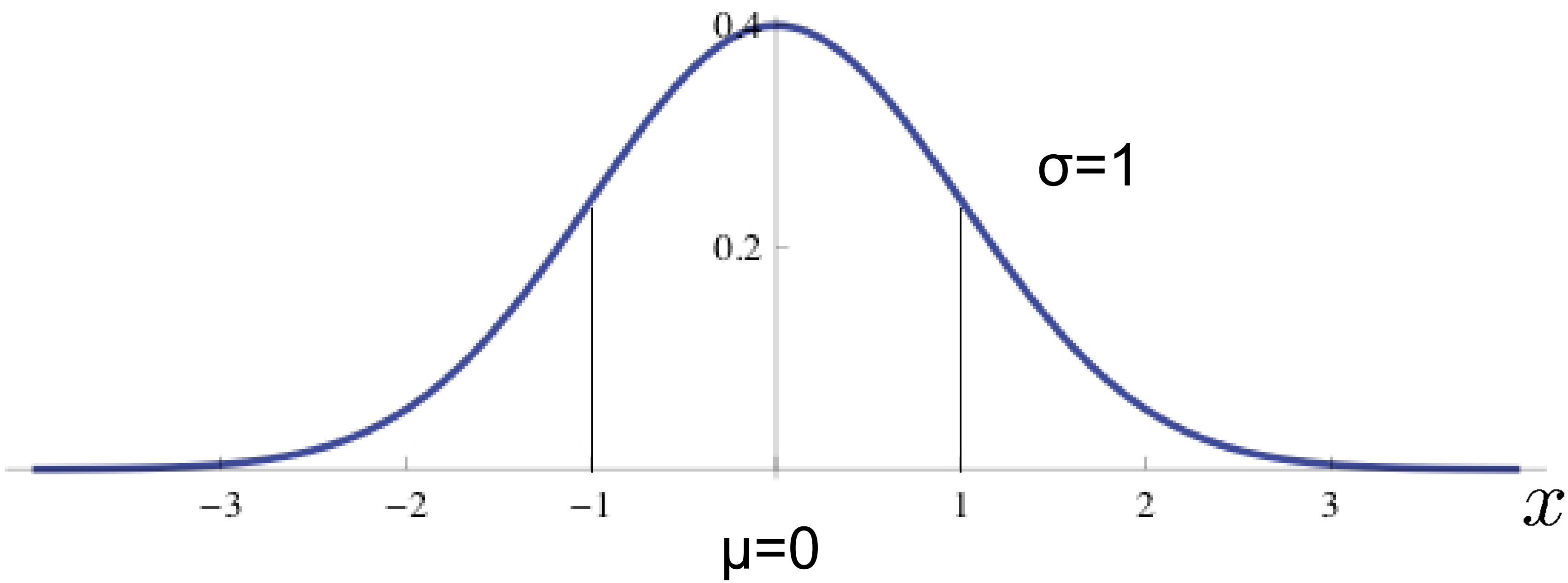
μ : 平均

σ^2 : 分散

前回までの内容（標準正規分布）

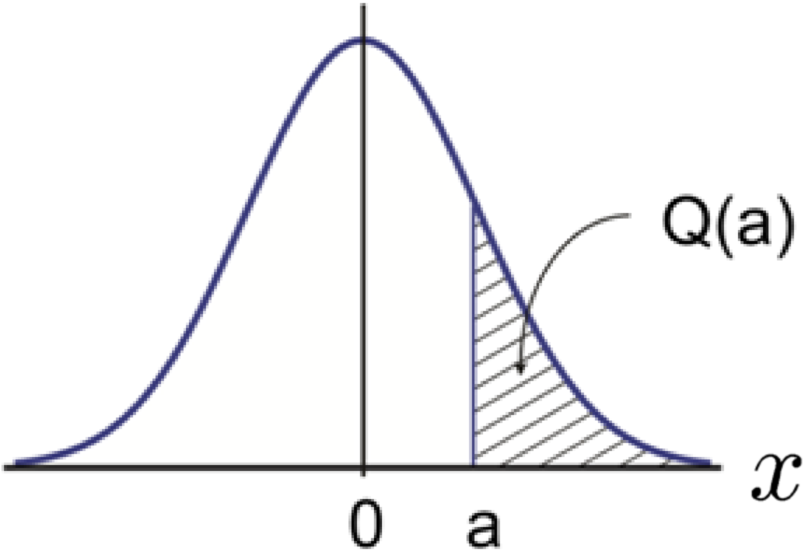
標準正規分布 $\mathrm{N}(0,1)$ の確率密度関数

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty < x < \infty)$$

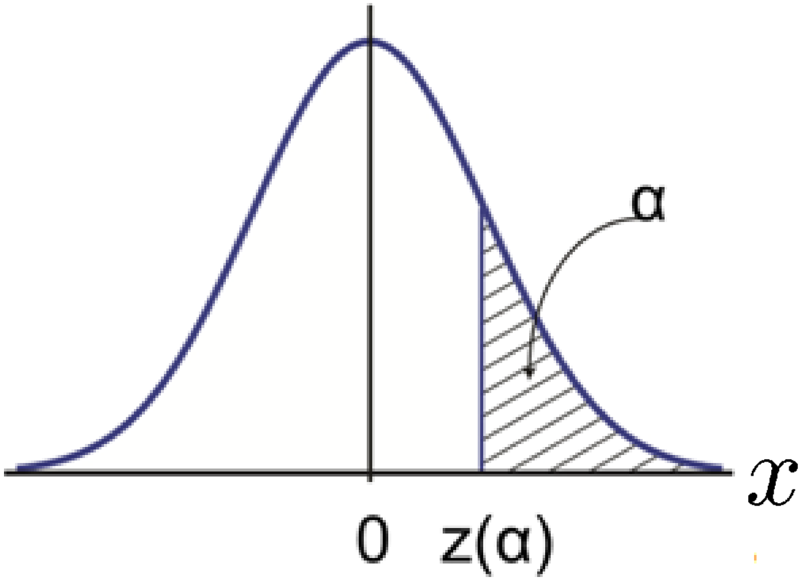


前回の復習（標準正規分布）

・上側確率：$Q(a) = \Pr(X \geq a)$



・上側 α 点：$\alpha = \Pr(X \geq x)$ を満たす x を $z(\alpha)$ と表す.



例　2．19

歪みのないサイコロを 500 回投げるとき，1 の目が出る回数が 80 回以上 100 回以下となる確率はいくらぐらいでしょう？

確率変数 X を 1 の目が出る回数とすると，求めたい確率は，

$$\Pr(80 \leq X \leq 100)$$

$X \sim B\left(500, \frac{1}{6}\right)$ であるから，

$$\begin{aligned}\Pr(80 \leq X \leq 100) \quad = \quad & {}_{500}\mathrm{C}_{80}\left(\frac{1}{6}\right)^{80}\left(\frac{5}{6}\right)^{420} + {}_{500}\mathrm{C}_{81}\left(\frac{1}{6}\right)^{81}\left(\frac{5}{6}\right)^{419} \\ & + \cdots + {}_{500}\mathrm{C}_{100}\left(\frac{1}{6}\right)^{100}\left(\frac{5}{6}\right)^{400}\end{aligned}$$

注意

${}_{500}\mathrm{C}_{80}, {}_{500}\mathrm{C}_{81}, \ldots, {}_{500}\mathrm{C}_{100}$ の計算は大変である．

例　2. 20

日本における男女の出生比率はおよそ 51 : 49 です．いま，1 万人の子供が生まれたとすると，そのうち男児の総数は直観的には 5100 人程度と予想されます．それでは，男児の総数が 5050 人以上 5150 人以下である確率はどの程度でしょうか．

確率変数 X を生まれた男児の総数とすると，$X \sim \mathrm{B}(10000, 0.51)$

$$\Pr(5050 \leq X \leq 5150) = {}_{10000}\mathrm{C}_{5050}\ 0.51^{5050}\ 0.49^{4950}$$
$$+ {}_{10000}\mathrm{C}_{5051}\ 0.51^{5051}\ 0.49^{4949}$$
$$+ \cdots + {}_{10000}\mathrm{C}_{5150} 0.51^{5150}\ 0.49^{4850}$$

⇩

2 つの例に対して，正規近似を用いて確率を計算する．

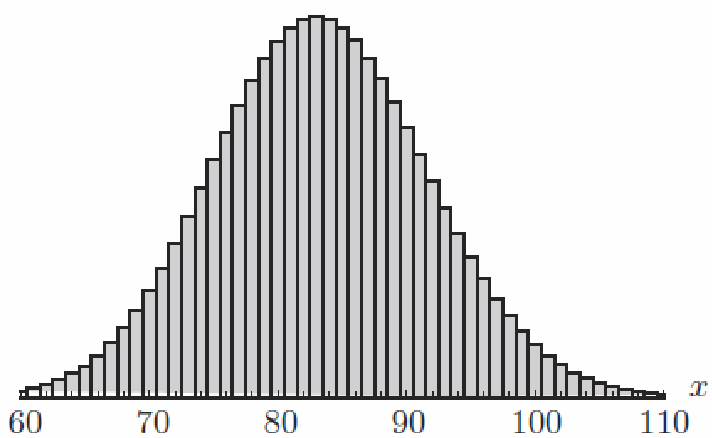


図 2.32　2 項分布 $\mathrm{B}(500, \frac{1}{6})$ の確率関数のグラフ（例 2.19）

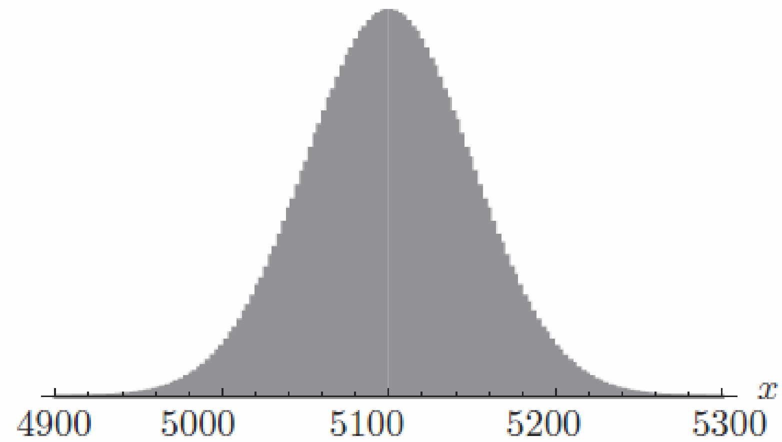
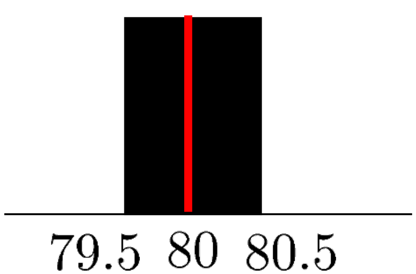


図 2.33　2 項分布 $\mathrm{B}(10000, 0.51)$ の確率関数のグラフ（例 2.20）

・注意

確率関数には, 幅 1 の長方形の面積を対応させている.



$\Pr(X = 80)$ (赤い線) を 79.5 から 80.5 の長方形の面積で表している.

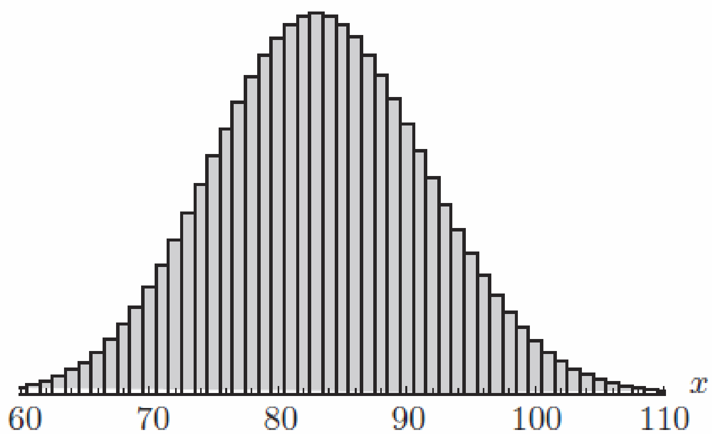


図 2.32　2 項分布 $B(500, \frac{1}{6})$ の確率関数のグラフ（例 2.19）

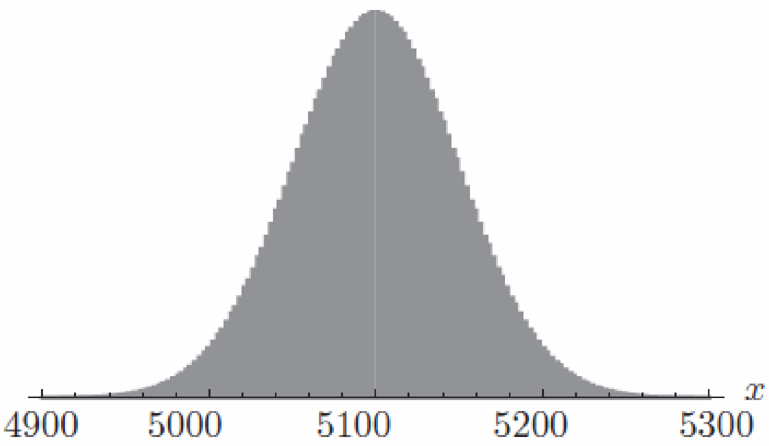


図 2.33　2 項分布 $B(10000, 0.51)$ の確率関数のグラフ（例 2.20）

これらの確率関数は正規分布の確率密度関数で近似される.
つまり, 1 の目が出る回数が 80 回以上 100 回以下となるおよその確率や, 男児の総数が 5050 人以上 5150 人以下であるおよその確率は正規分布表（数表1）を用いて求めることが出来る.

公式 2.10.(中心極限定理)

確率変数 X は２項分布 $\mathrm{B}(n,p)$ に従っているものとする．このとき，n が十分大きいならば，確率変数 X は近似的に正規分布 $\mathrm{N}(np, np(1-p))$ に従う．つまり，

$$\frac{X-np}{\sqrt{np(1-p)}}$$

は近似的に標準正規分布 $\mathrm{N}(0,1)$ に従う．

注意 2.4

公式 2.10 の「X は近似的に正規分布に従う」という意味は２項分布のおよその確率が正規分布の確率で求められることを表している．このことを２項分布の**正規近似**という．

実用上適用可能な精度の近似を与える条件

特に，公式 2.10 の設定で，実用上適用可能な精度の近似を与える条件として

$$np \geq 5 \quad \text{かつ} \quad n(1-p) \geq 5$$

が知られている．
たとえば $p = 0.5$ であれば $n \geq 10$ で実用上適用可能になる．

例 2.19 の場合 $(n = 500, p = 1/6)$ を考える．

$$np = 500 \times \frac{1}{6} = 83.33\cdots \geq 5, \quad n(1-p) = 500 \times \frac{5}{6} = 416.6\cdots \geq 5$$

となるので，正規近似は十分使える．

$$\Pr(80 \leq X \leq 100) \quad \Pr(100 \leq X)$$

$$\Pr(X \leq 100) \quad \Pr(X = 100)$$

また，公式 2.10 より

$$Z = \frac{X - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}}$$

は近似的に標準正規分布 N(0, 1) に従う．

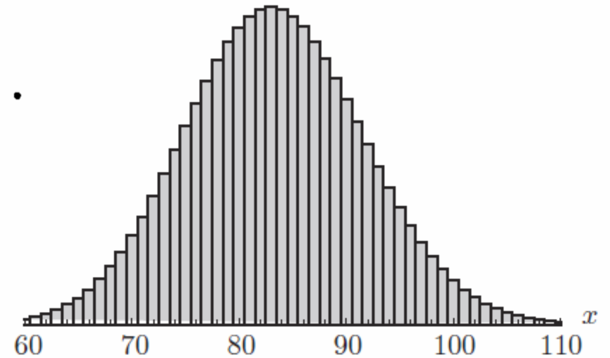


図 2.32　2 項分布 $B(500, \frac{1}{6})$ の確率関数のグラフ（例 2.19）

実際に確率を計算する際には，
図 2.32 と同じように，

$$\Pr(80 \le X \le 100) = \Pr(80 - 0.5 \le X \le 100 + 0.5)$$

とする．

このように 0.5 を加減することを**連続修正**という．

図 2.34 は 2 項分布 $\mathrm{B}(500, \frac{1}{6})$ の確率関数（太線）と
正規分布 $\mathrm{N}(500 \times \frac{1}{6}, 500 \times \frac{1}{6} \times \frac{5}{6})$ の確率密度関数（曲線）である．

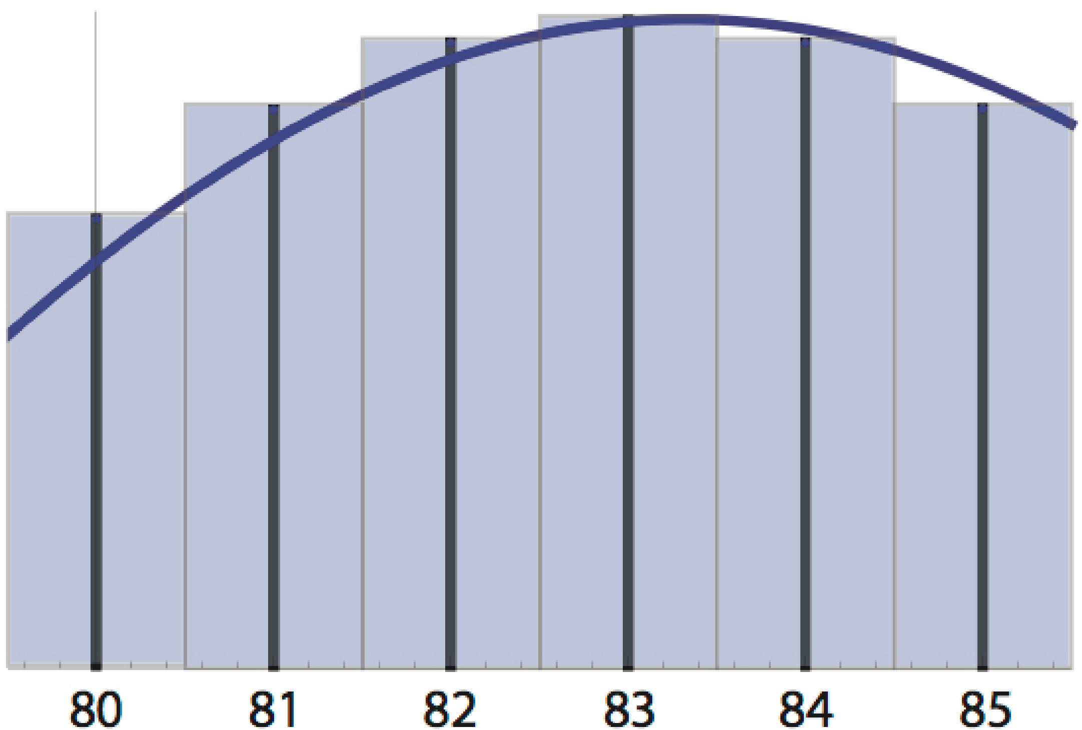


図 **2.34**　確率関数と確率密度関数のグラフ（例 **2.19**）

これらのことから

$$
\begin{aligned}
& \Pr(80 \leq X \leq 100) \\
= \; & \Pr(80 - 0.5 \leq X \leq 100 + 0.5) \\
= \; & \Pr\left(\frac{80 - 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \leq \frac{X - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \leq \frac{100 + 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}}\right) \\
= \; & \Pr\left(-0.46 \leq Z \leq 2.06\right) \\
= \; & Q(-0.46) - Q(2.06)
\end{aligned}
$$

となる.

これらのことから

$$\Pr(80 \leq X \leq 100) = Q(-0.46) - Q(2.06) \quad (2.17)$$

となる．

ここで公式 2.8 (1), (2) から

$$Q(-0.46) = 1 - Q(0.46)$$

なので，(2.17) より

$$\Pr(80 \leq X \leq 100) = 1 - Q(0.46) - Q(2.06) = 0.6575$$

となる．

つまり，1 の目が出る回数が 80 回以上 100 回以下となる確率は 0.66 程度である．

ちなみに,

$$\begin{aligned}\Pr(80 \le X \le 100) &= {}_{500}\mathrm{C}_{80}\left(\frac{1}{6}\right)^{80}\left(\frac{5}{6}\right)^{420} + {}_{500}\mathrm{C}_{81}\left(\frac{1}{6}\right)^{81}\left(\frac{5}{6}\right)^{419} \\ &\quad + \cdots + {}_{500}\mathrm{C}_{100}\left(\frac{1}{6}\right)^{100}\left(\frac{5}{6}\right)^{400}\end{aligned}$$

の右辺を計算ソフトを用いて直接計算すると

$$\Pr(80 \le X \le 100) = 0.6518\cdots$$

となる. 近似値 (0.6575) と比べてみると, 小数第 2 位まで一致していることがわかる.

他の計算例

(i)
$$
\begin{aligned}
\Pr(100 \le X) &= \Pr(100 - 0.5 \le X) \\
&= \Pr\left(\frac{100 - 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \le \frac{X - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}}\right) \\
&= \Pr(1.94 \le Z) = Q(1.94) = 0.02619
\end{aligned}
$$

$\Pr(100 \le X)$ を正規近似で求める場合には，$\Pr(100 \le X \le 500)$ としないこと．

(ii)
$$
\begin{aligned}
\Pr(X \le 100) &= \Pr(X \le 100 + 0.5) \\
&= \Pr\left(\frac{X - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \le \frac{100 + 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}}\right) \\
&= \Pr(Z \le 2.06) = 1 - \Pr(Z > 2.06) \\
&= 1 - Q(2.06) = 0.9803
\end{aligned}
$$

$\Pr(X \le 100)$ を正規近似で求める場合には，$\Pr(0 \le X \le 100)$ としないこと．

他の計算例

(iii)

$$
\begin{aligned}
\Pr(X = 100) &= \Pr(100 - 0.5 \le X \le 100 + 0.5) \\
&= \Pr\left(\frac{100 - 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \le \frac{X - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}} \le \frac{100 + 0.5 - 500 \times \frac{1}{6}}{\sqrt{500 \times \frac{1}{6} \times \frac{5}{6}}}\right) \\
&= \Pr\left(1.94 \le Z \le 2.06\right) = Q(1.94) - Q(2.06) = 0.00649
\end{aligned}
$$

例　2．20

日本における男女の出生比率はおよそ $51:49$ です．いま，1 万人の子供が生まれたとすると，そのうち男児の総数は直観的には 5100 人程度と予想されます．それでは，男児の総数が 5050 人以上 5150 人以下である確率はどの程度でしょうか．

$$np = 10000 \times 0.51 = 5100 \geq 5, \quad n(1-p) = 10000 \times 0.49 = 4900 \geq 5$$

となるので，正規近似は適応可能．

公式 2.10 より,

$$Z = \frac{X - 10000 \times 0.51}{\sqrt{10000 \times 0.51 \times 0.49}}$$

は近似的に標準正規分布 N(0,1) に従う. このことから

例 2.20

このことから

$$
\begin{aligned}
& \Pr(5050 \leq X \leq 5150) \\
= \; & \Pr\left(5050 - 0.5 \leq X \leq 5150 + 0.5\right) \\
= \; & \Pr\left(\frac{5050 - 0.5 - 10000 \times 0.51}{\sqrt{10000 \times 0.51 \times 0.49}} \leq \frac{X - 10000 \times 0.51}{\sqrt{10000 \times 0.51 \times 0.49}}\right. \\
& \qquad \left. \leq \frac{5150 + 0.5 - 10000 \times 0.51}{\sqrt{10000 \times 0.51 \times 0.49}}\right) \\
\fallingdotseq \; & \Pr\left(-1.01 \leq Z \leq 1.01\right) = Q(-1.01) - Q(1.01) \\
= \; & 1 - Q(1.01) - Q(1.01) = 0.6876
\end{aligned}
$$

となる．

正確な値は $\Pr(5050 \leq X \leq 5150) = 0.6876 \cdots$

２．１３　母集団と標本

・第 1 章では，与えられた標本をどのように整理するか，また，標本平均，標本分散や不偏分散等の代表値を用いて，標本の中心的位置やばらつきのような特徴を見出すことを考えた.

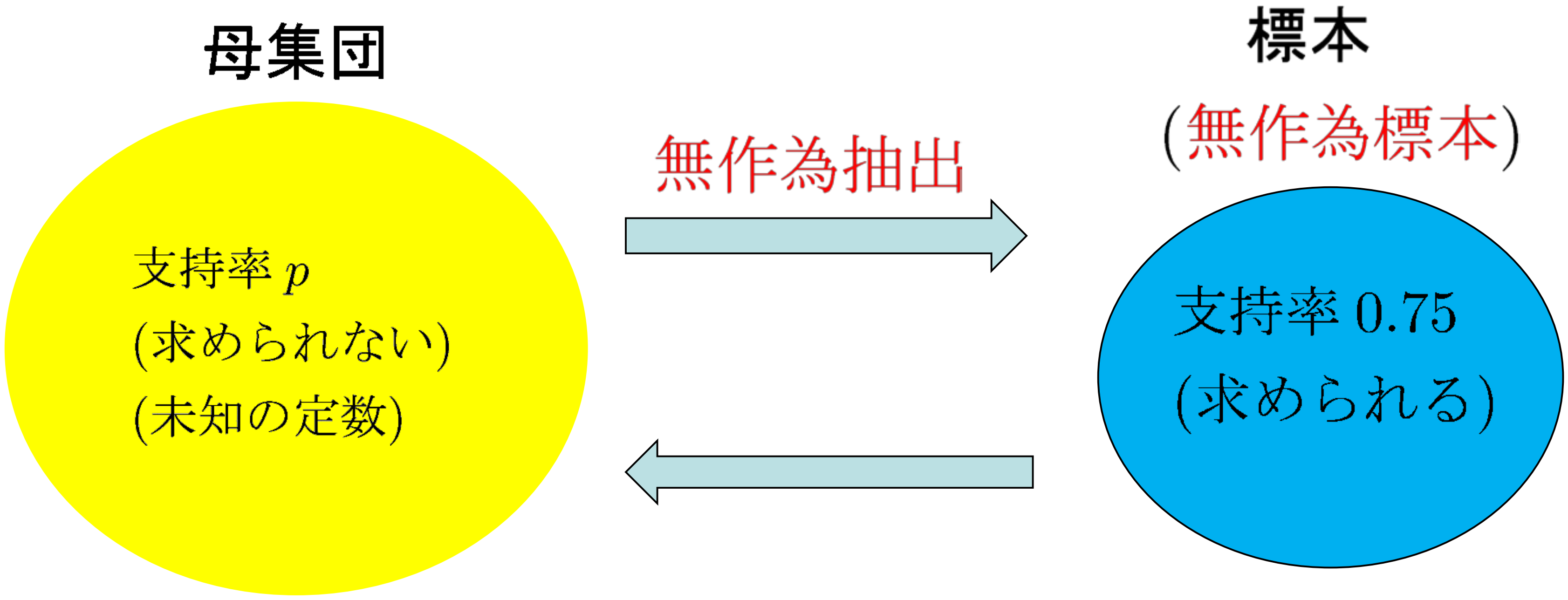
この節では，標本と確率変数，確率分布の関係について考える.

支持率調査

- 日本の首相の支持率調査

    母集団：有権者全体 (調査対象の全体)

    標本：選ばれた 1000 人の有権者 (母集団の一部)



母集団の支持率（母比率）を推測

無作為抽出法と無作為標本

母集団から標本を選ぶとき，**偏った有権者だけを選んではいけない.**
母集団の様子をうまく反映するために，どの有権者も同等に選ばれる必要がある.
つまり，母集団のどの有権者も等しい確率で選ぶことになる．このような標本の選び方を**無作為抽出法**といい，選ばれた標本を**無作為標本**という.

母集団での支持率と標本での支持率

- **母集団での支持率の推測:**

  母集団での支持率を p とし，標本での支持率を 75% とすると，
  p は 0.75 ぐらいと推測される．

- 1000 **人の支持率調査** (5 **回**)**:**

$$75\%, \quad 72\%, \quad 68\%, \quad 73\%, \quad 80\%$$

母集団の支持率 p は未知の定数であるが，標本での支持率は変数であり，確率変数であると考えられる．その確率変数の値（この値を**実現値**という）が 0.75, 0.72, 0.68, 0.73, 0.80 である．

標本での支持率が確率変数の実現値であると考えられる理由

・**調査結果を数値化する**：

支持することを 1, 支持しないことを 0 と表す.

・1 **回目の支持率調査の結果**:

$$1, \quad 0, \quad 0, \quad 1, \quad \cdots, \quad 0, \quad 1, \quad 1, \quad 0$$

(1 の個数が 750, 0 の個数が 250)

最初の有権者は 1 であるので，支持する有権者が選ばれたことになるが，どの有権者も等しい確率で選ばれることから，最初の有権者の 1 は

$$\Pr(X = 1) = p, \quad \Pr(X = 0) = 1 - p$$

となる確率変数 X（２項分布 $\mathrm{B}(1, p)$ に従う）の実現値と考えられる.

標本での支持率が確率変数の実現値であると考えられる理由

1 回目の支持率調査の結果:

$$1, \quad \boxed{0, \quad 0, \quad 1, \quad \cdots, \quad 0, \quad 1, \quad 1, \quad 0}$$

(1 の個数が 750, 0 の個数が 250)

2 人目から 1000 人目の調査結果も 2 項分布 $\mathrm{B}(1, p)$ に従う確率変数の実現値と考えられる.

標本での支持率

$$\frac{\text{支持する有権者の人数}}{1000} = 0.75$$

も確率変数の実現値と考えられる.

標本での支持率と母集団での支持率

母集団からの無作為標本：

$$X_1, X_2, \cdots, X_n$$

が２項分布 $\mathrm{B}(1, p)$ に従う確率変数とするとき，標本での支持率はこれらの標本平均：

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{1000}$$

と表される.

標本での支持率		母集団での支持率
$\bar{X}$ の実現値	推測 →	p

実際には，１回の支持率調査の実現値から p を推測することになるが，

$$X_1, X_2, \cdots, X_{1000}$$

は確率変数であることより，**標本での支持率 $\bar{X}$ がどのようにばらつくかが予測できる**.

母集団分布と母比率

無作為標本	X_1	X_2	X_3	X_4	$\cdots$	X_{997}	X_{998}	X_{999}	X_{1000}	$\bar{X}$
実現値	1	0	0	1	$\cdots$	0	1	1	0	0.75

$X_1, X_2, \cdots, X_{1000}$ が 2 項分布 $\mathrm{B}(1, p)$ に従うとして議論してきた.

この 2 項分布 $\mathrm{B}(1, p)$ のように母集団に想定される確率分布を**母集団分布**といい，母集団での比率 (図 2.35 では支持率) p を**母比率**という.

標本調査法

- **標本調査法**

母集団から標本を抽出して，調査する方法.

- **全数調査法**

母集団全体を調査する方法.

5 年に 1 度行われる国勢調査

ある製品の耐久力試験では，すべての製品に試験を行った場合，試験後にはまともな製品は一つも残らないことになり，この場合には全数調査法は不可能.

学力調査（全数調査から標本調査に変更）

例 2.21

子供の学力状況を把握するため，文部科学省は 2007 年に全国学力・学習状況調査を行った．

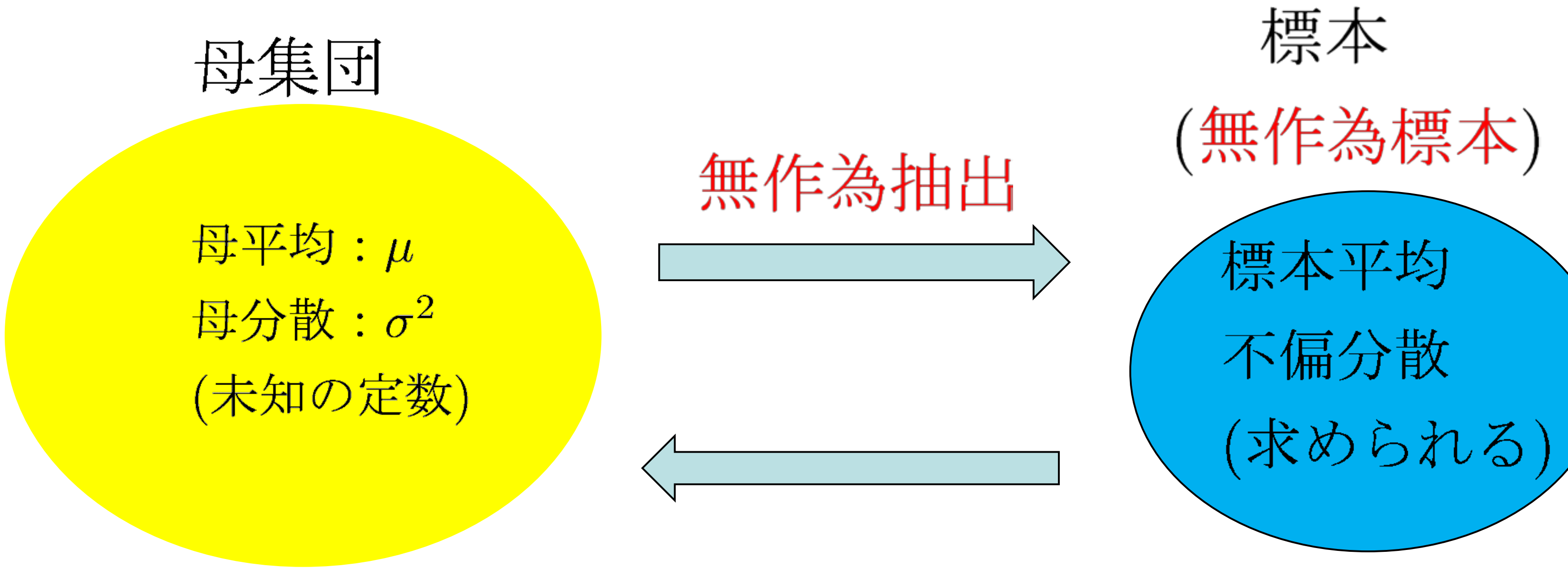
これは基本的に小学 6 年生と中学 3 年生全員を対象とした全数調査を前提としたものであったが，諸々の理由で一部の小学生，中学生が受験しなかった．

全数調査では予算がかかりすぎるので，2010 年は基本的に標本調査に変更となった．

標本調査(20 才の日本人男性の身長の平均と分散)

母集団：20 才の日本人男性の全体

母集団分布に, 正規分布 $\mathrm{N}(\mu, \sigma^2)$ を想定する.



母集団の平均（母平均）と分散（母分散）を推測する

3章以降

第３章以降では，標本がどのように与えられるかを考え，母集団分布を想定し，標本を母集団からの無作為標本の実現値と捉えることによって，母集団に関する推測を行います．このような考え方が第３章の推定，第４章の検定等でみられる結果の基礎となっています．なお，母比率，母平均，母分散のように母集団分布を特徴づける特性値を総称して**母数**といいます．

また，母集団分布が２項分布の母集団を**２項母集団**といい，母集団分布が正規分布の母集団を**正規母集団**といいます．

まとめ

中心極限定理

２項分布の正規近似

連続修正

母集団、標本、無作為抽出、無作為標本

母集団分布、母比率、母平均、母分散

標本平均、標本分散、不偏分散