

統計学の整理帳

tomixy

2025年8月7日

目次

第 1 章	一次元データの代表値と散らばり	3
	データの中心の指標：平均値	3
	データのばらつきの指標：偏差	3
	データのばらつきの指標：分散と標準偏差	5
	分散公式	6
	データの変換による平均と分散の変化	6
	データの標準化	8
	度数分布	9
	標本平均	11
第 2 章	二次元データの相関	12
	変量の相互関係：相関	12
	相関の数値化：共分散	13
第 3 章	確率の基本	15
	論理から確率へ	15

出来事の結果と確率の計算	16
場合の数：和の法則と積の法則	18
和事象の確率	21
排反な事象と和事象の確率	22
条件つき確率	23
積事象の確率	23
独立な事象と積事象の確率	24
第 4 章 確率分布	26
記述統計から推測統計へ	26
母集団からの無作為抽出	26
母集団の度数分布	27
相対度数と確率	28
標本と確率変数	29
母集団の度数分布と確率分布	29
連続型確率分布	30
離散型確率分布	33

第 1 章

一次元データの代表値と散らばり

データの中心の指標：平均値

「データを 1 つの値で要約するならばこれ」といった指標を代表値という。

最もよく使われる代表値が平均値 (mean) であり、データの中心を表す指標として広く用いられる。

平均値は、データをすべて足し合わせて、データの数で割ることで求まる。

 平均 N 個の観測値 x_1, \dots, x_N の総和をデータのサイズ N で割ったものを平均値という。

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$$

データのばらつきの指標：偏差

代表値はデータを 1 つの値で要約する指標であり、データのばらつきや偏りは表現しきれない。そこで、新たにデータのばらつきを表す指標を考える。

各データが、平均からどれくらい離れているかを表す指標を**偏差 (deviation)** という。

 **偏差** N 個の観測値 x_1, \dots, x_N の平均値を \bar{x} とするとき、各観測値 x_i の**偏差**は次のように定義される。

$$d_i := x_i - \bar{x}$$

ここで、 d_i は i 番目のデータの偏差を表す。

偏差の平均値で全体をみる

全データの偏差 d_1, \dots, d_N の平均値を求めることで、データ全体が平均からどれくらい離れて分布しているか（どれくらいばらついているか）を表すことができそうである。

しかし、偏差の平均値は、次のように常に 0 になってしまう。

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N d_i &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) \\ &= \frac{1}{N} \left(\sum_{i=1}^N x_i - \sum_{i=1}^N \bar{x} \right) = \frac{1}{N} \left(\sum_{i=1}^N x_i - N\bar{x} \right) \\ &= \sum_{i=1}^N \frac{1}{N} x_i - \bar{x} = \bar{x} - \bar{x} = 0 \end{aligned}$$

そこで、単なる平均との差ではなく、平均との距離を考えることにする。

偏差に絶対値をつけたものの平均を**平均偏差**という。

 **平均偏差** N 個の観測値 x_1, \dots, x_N の平均値を \bar{x} とするとき、**平均偏差**を次のように定義する。

$$d := \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|$$



データのばらつきの指標：分散と標準偏差

平均偏差では、データと平均値の距離として絶対値を用いたが、絶対値は次のような理由で計算が面倒である。

- 絶対値は微分できない点がある
- 正負を判定する条件分岐処理が入り、コンピュータでの計算速度が落ちる

そこで、絶対値の代わりに二乗を用いた、**分散 (variance)** という指標を定義する。

 **分散** N 個の観測値 x_1, \dots, x_N の平均値を \bar{x} とするとき、**分散** を次のように定義する。

$$\sigma^2 := \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

もとのデータと同じ単位を持ったばらつきの指標

分散では二乗を用いるため、単位に注意が必要である。

もとのデータの単位が $[x]$ であれば、分散の単位は $[x]^2$ となる。

たとえば、点数を表すデータを扱っているとすると、その分散の単位は「点²」となり、直観的に理解しづらい。

そこで、単位をもとのデータと揃えるために、分散の平方根をとった形がよく用いられる。

分散の平方根を**標準偏差 (standard deviation)** という。

 **標準偏差** 分散 σ^2 の平方根をとったものを**標準偏差**として定義する。

$$\sigma := \sqrt{\sigma^2}$$


分散公式

分散は、次のように計算することもできる。

分散 = データの二乗平均 - 平均の二乗

 分散公式 N 個の観測値 x_1, \dots, x_N の平均を \bar{x} 、分散を σ^2 とすると、次の関係が成り立つ。

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2$$

証明

分散の定義に基づいて、次のように計算する。

$$\begin{aligned}\sigma^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \\ &= \frac{1}{N} \sum_{i=1}^N (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{N} \sum_{i=1}^N x_i^2 - 2\bar{x} \frac{1}{N} \sum_{i=1}^N x_i + \bar{x}^2 \\ &= \frac{1}{N} \sum_{i=1}^N x_i^2 - 2\bar{x}^2 + \bar{x}^2 \\ &= \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2\end{aligned}$$

データの変換による平均と分散の変化

データの変換（スケーリングやシフト）を行うと、平均や分散はどのように変化するのだろうか。

データのスケールリング

まず、データを定数 a 倍する変換を考える。

すなわち、各データ x_i を $y_i = ax_i$ に変換すると、平均と分散は次のように変化する。

$$\begin{aligned}\bar{y} &= a\bar{x} \\ \sigma_y^2 &= a^2\sigma_x^2\end{aligned}$$

 データのスケールリングによる平均と分散の変化 観測値が a 倍されると、平均は a 倍、分散は a^2 倍される。

証明

各データ x_i を $y_i = ax_i$ に変換すると、平均は次のように変化する。

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N ay_i = a \cdot \frac{1}{N} \sum_{i=1}^N y_i = a\bar{y}$$

また、分散は次のように変化する。

$$\begin{aligned}\sigma_y^2 &= \frac{1}{N} \sum_{i=1}^N (ax_i - a\bar{x})^2 = \frac{1}{N} \sum_{i=1}^N (a(x_i - \bar{x}))^2 \\ &= a^2 \cdot \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = a^2\sigma_x^2\end{aligned}$$

データのシフト

次に、データを定数 b だけシフトする変換を考える。

すなわち、各データ x_i を $y_i = x_i + b$ に変換すると、平均と分散は次のように変化する。

$$\begin{aligned}\bar{y} &= \bar{x} + b \\ \sigma_y^2 &= \sigma_x^2\end{aligned}$$

 データのシフトによる平均と分散の変化 観測値に b を加えると、平均は b だけ増えるが、分散は変化しない。

証明

各データ x_i を $y_i = x_i + b$ に変換すると、平均は次のように変化する。

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N (x_i + b) = \frac{1}{N} \sum_{i=1}^N x_i + b = \bar{x} + b$$

また、分散は次のように変化しない。

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (x_i + b - (\bar{x} + b))^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \sigma_x^2$$

このように、データの変換によって平均と分散は異なる影響を受けることがわかる。

データの標準化

たとえば、平均点が 30 点のテストでとった 60 点と、平均点が 80 点のテストでとった 60 点とでは、相対的な出来が異なる。このように、点数というデータはそのテストの平均や分散によって評価が変わってしまう。

そのため、平均や分散に依存せずにデータの相対的な位置関係がわかるようにできたら便利である。

特に、平均が 0、標準偏差が 1 になるようにデータを変換することを**標準化 (standardization)**という。

$y_i = ax_i + b$ というデータの変換を考えよう。

これは、データを a 倍して b だけシフトする変換である。

このとき、平均と標準偏差は次のように変化する。

- 平均は a 倍され、 b だけ増える
- 標準偏差は a 倍される（分散が a^2 倍される）

数式で表すと、

$$\begin{aligned}\bar{y} &= a\bar{x} + b \\ \sigma_y &= a\sigma_x\end{aligned}$$

そこで、平均 \bar{y} が 0、標準偏差 σ_y が 1 になるように、 a と b を次のように設定する。

$$a = \frac{1}{\sigma_x}, \quad b = -\frac{\bar{x}}{\sigma_x}$$

このようにすると、たしかに $\bar{y} = 0$ 、 $\sigma_y = 1$ となる。

このとき、変換後のデータ y_i は次のように表される。

$$y_i = ax_i + b = \frac{x_i - \bar{x}}{\sigma_x}$$

 **標準化** 各データから平均を引き、標準偏差で割ることで、平均が 0、標準偏差が 1 になるように変換することを**標準化**という。

各データ x_i を標準化したデータを y_i とすると、次の関係が成り立つ。

$$y_i = \frac{x_i - \bar{x}}{\sigma_x}$$

度数分布

より細かくデータがどのように分布しているかを知りたいときは、次のような手順でデータを整理する。

1. データがとる値をいくつかの区間に分ける
2. 各区間にいくつかのデータが入っているかを数える

このとき、区間を**階級 (class)** といい、各階級に属しているデータの数を**度数 (frequency)** という。

各階級ごとに、度数などの値を表にまとめたものを**度数分布表**という。

度数分布表

階級	階級値	度数	累積度数	相対度数	累積相対度数
$a_0 \sim a_1$	x_1	f_1	F_1	$\frac{f_1}{N}$	$\frac{F_1}{N}$
$a_1 \sim a_2$	x_2	f_2	F_2	$\frac{f_2}{N}$	$\frac{F_2}{N}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$a_{n-1} \sim a_n$	x_n	f_n	F_n	$\frac{f_n}{N}$	$\frac{F_n}{N}$

階級値

各階級を代表する値を**階級値**といい、通常は各階級の中央の値を階級値として用いる。

$$x_i = \frac{a_{i-1} + a_i}{2}$$

累積度数

その階級までに属するデータの個数を**累積度数**という。

$$F_i = \sum_{j=1}^i f_j$$

相対度数

全データ数に対して、その階級に属するデータ数がどのくらいの割合を占めているのかを**相対度数**という。

$$\frac{f_i}{N}$$

累積相対度数

全データ数に対して、その階級までに属するデータ数がどのくらいの割合を占めているのかを**累積相対度数**という。

$$\frac{F_i}{N}$$



標本平均

度数分布から求めた平均値を**標本平均**という。

度数分布表

階級値	x_1	x_2	...	x_n
度数	f_1	f_2	...	f_n

上のような度数分布表が与えられたとき、1つの階級に属するデータ数は f_i 個であり、これらがすべて同じ値 x_i をとると考えて平均を求める。

全データ数を N とすると、

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i$$

これは通常の平均とは異なり、すべてのデータを使っているわけではないことを強調するため、**標本平均**と呼ばれる。

 **標本平均** i 番目の階級の階級値を x_i 、度数を f_i とし、全データ数を N とするとき、**標本平均**を次のように定める。

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n f_i x_i$$

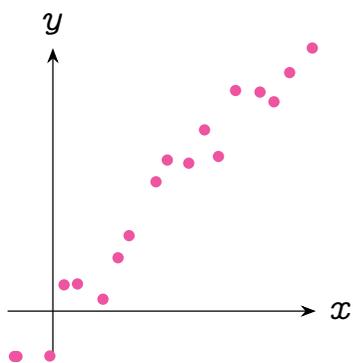
第 2 章

二次元データの相関

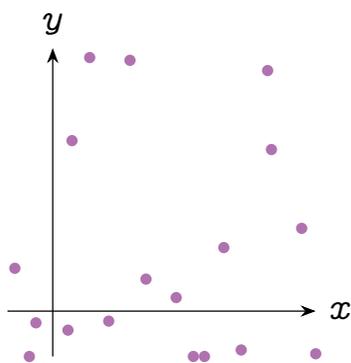
変量の相互関係：相関

2次元データ (x, y) において、2つの変量 x と y の間に相互関係がみられるとき、 x と y の間には相関 (correlation) 関係があるという。

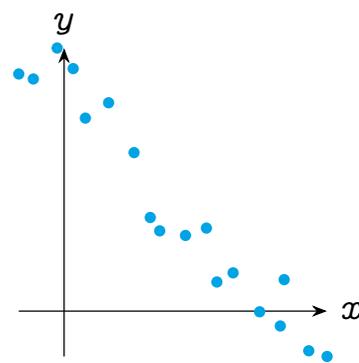
- **正の相関** : x が大きいほど、 y も大きくなる傾向がある
- **負の相関** : x が大きいほど、 y は小さくなる傾向がある
- **無相関** : どちらにも当てはまらない (直線的な関係がない)



正の相関



無相関

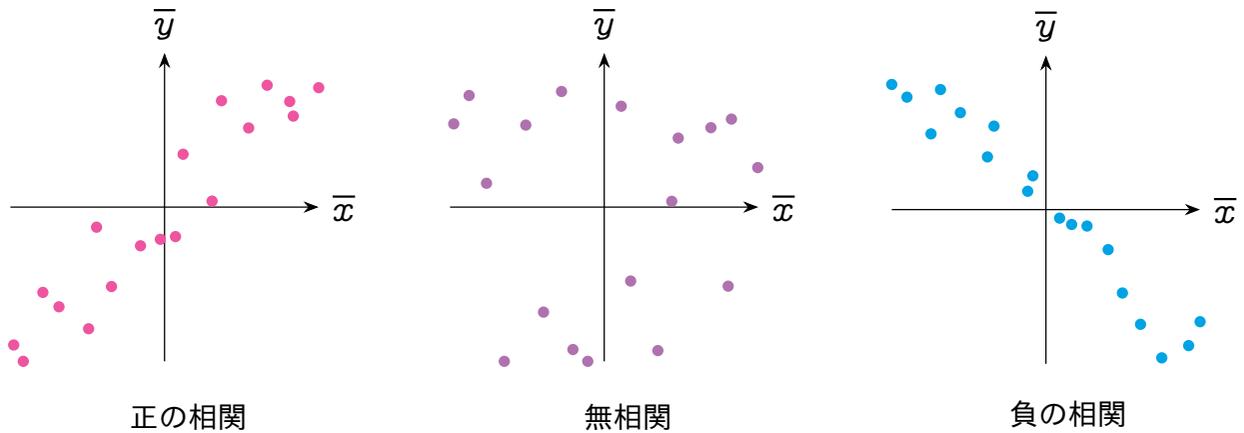


負の相関

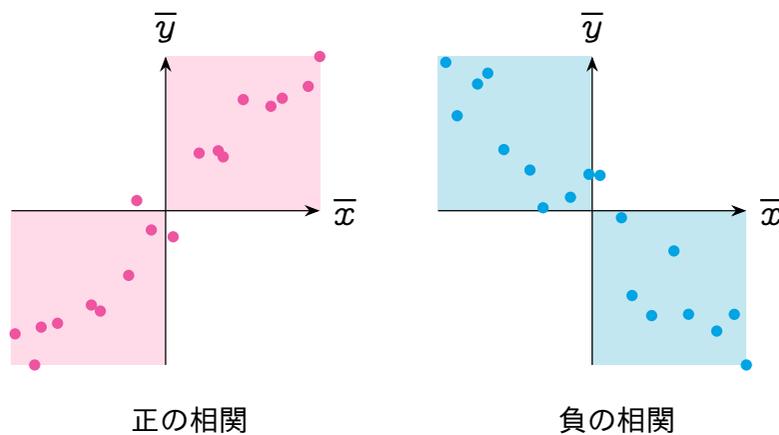
相関の数値化：共分散

グラフを描いて視覚的に相関を確認することはできるが、客観的に表現するために、数値で判断できるようにしたい。

そのために、 x, y の平均 (\bar{x}, \bar{y}) を原点とする新たな座標軸を考える。



すると、正の相関か負の相関かに応じて、データが多く分布する象限（座標軸で切り分けた領域）が異なることがわかる。



正の相関の場合は、第一象限と第三象限にデータが多く分布することがわかる。

- 第一象限： $x > \bar{x}$ かつ $y > \bar{y}$ である範囲
- 第三象限： $x < \bar{x}$ かつ $y < \bar{y}$ である範囲

負の相関の場合は、第二象限と第四象限にデータが多く分布することがわかる。

- 第二象限： $x < \bar{x}$ かつ $y > \bar{y}$ である範囲
- 第四象限： $x > \bar{x}$ かつ $y < \bar{y}$ である範囲

この場合分けは、次のようにまとめることができる。

- 正の相関の場合、 $x - \bar{x}$ と $y - \bar{y}$ の符号が同じになる点が多い
- 負の相関の場合、 $x - \bar{x}$ と $y - \bar{y}$ の符号が反対になる点が多い

さらに、符号が同じものの積は正、符号が反対のもの積は負になることから、

- 正の相関の場合、 $(x - \bar{x})(y - \bar{y}) > 0$ となる点が多い
- 負の相関の場合、 $(x - \bar{x})(y - \bar{y}) < 0$ となる点が多い

各データについて $(x_i - \bar{x})(y_i - \bar{y})$ を求め、全データの平均をとることで、相関を判定できそうである。このような考え方で相関を数値化したものを**共分散 (covariance)** という。

 **共分散** N 個の観測値 $(x_1, y_1), \dots, (x_N, y_N)$ の平均をそれぞれ \bar{x}, \bar{y} とするとき、**共分散** を次のように定義する。

$$\sigma_{xy} := \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

第 3 章

確率の基本

論理から確率へ

確率 (probability) は、論理を拡張したものと捉えることができる。

論理では真と偽という 2 つの値があり、これらは確信 (絶対的な信念) に対応する。
なにかが真であるというのは、それが「正しい」と完全に確信しているという意味である。

しかし、私たちが行う決定には、ほぼ必ず、確信のなさがある程度伴っている。
確率を使えば、論理を拡張して、「真と偽の間の不確実な値」を扱うことができる。

確率における真と偽

真は 1、偽は 0 で表現することが多いので、確率の定義もそれに倣うことにする。

X である確率を $P(X)$ とすると、

- $P(X) = 0$: X が偽
- $P(X) = 1$: X が真

0 と 1 の間には無限個の数が存在し、どちらの確信の方が強いかによってこの値が揺らぐ。

- 0 の方に近い値は、ある事柄 X が偽である確信の方が強いという意味
- 1 の方に近い値は、ある事柄 X が真である確信の方が強いという意味

- 0.5 という値は、ある事柄 X の真偽にまったく確信が持てないという意味

確率における否定

論理で重要なものとして、**否定**がある。

- 「真でない」とは「偽である」という意味
- 「偽でない」とは「真である」という意味

確率にもこのような性質を与えたいので、 X である確率と X でない確率を足すと 1 になるようにする。

$$P(X) + P(\neg X) = 1$$

ここで、記号 \neg は否定を表す。

この論理を使えば、 X でない確率を次のように表現できる。

$$P(\neg X) = 1 - P(X)$$

このとき、 $P(X) = 1$ であれば $P(\neg X) = 0$ となり、基本的な論理法則と合致する。

出来事の結果と確率の計算

確率を計算するための最も一般的な方法は、「出来事の結果」を数え上げるというものである。

ここで、いくつかの用語を定義しておこう。

- **標本空間**：ある出来事に対して起こりうるすべての結果の集まり
- **事象**：関心のある結果の集まり（標本空間の部分集合）

起こりうるすべての結果のうち、関心のある結果（今確率を求めたい対象）だけを取り出したものが事象なので、事象は標本空間の部分集合といえる。

例：コインを 1 回投げたら表が出る確率

コインを 1 回投げたとき、起こりうる結果は「表が出る」「裏が出る」の 2 通りである。

この 2 つの結果をまとめたものが標本空間であり、 Ω と表すことが多い。

$$\Omega = \{ \text{表}, \text{裏} \}$$

知りたいのは表が出る確率なので、事象を A とすると、

$$A = \{ \text{表} \}$$

事象 A はたしかに標本空間 Ω の部分集合になっている。

確率を最も馴染みのある考え方にとらえると、**確率**とはある事象が起こる可能性であり、

起こりうるすべての場合のうち、ある事象が起こる場合の**割合**



として計算できる。

X が何通りあるかを $n(X)$ と表すことにすると、表が出る確率は次のように計算できる。

$$P(\text{表}) = \frac{n(\{ \text{表} \})}{n(\{ \text{表}, \text{裏} \})} = \frac{1}{2}$$

 **確率（頻度論的立場）** 標本空間を Ω 、事象を A とすると、事象 A が起こる確率は次のように計算できる。

$$P(A) = \frac{n(A)}{n(\Omega)}$$

ここで注意が必要なのは、割り算は全体を「均等に」分けることを前提とした演算であることだ。標本空間に含まれるすべての場合の数で割ったものを確率とみなすには、

どの事象も同程度に起こりうる（**同様に確からしい**）



という仮定が必要になる。



場合の数：和の法則と積の法則

今のところ、 $n(\Omega)$ がわからない限り、確率を計算することはできない。

しかし、複雑な例になると、起こりうる結果を数え上げるのが難しくなる。

そこで登場するのが組み合わせ論（場合の数についての理論）である。

何通りの「場合」が起こり得るかを数え上げたものを**場合の数**という。

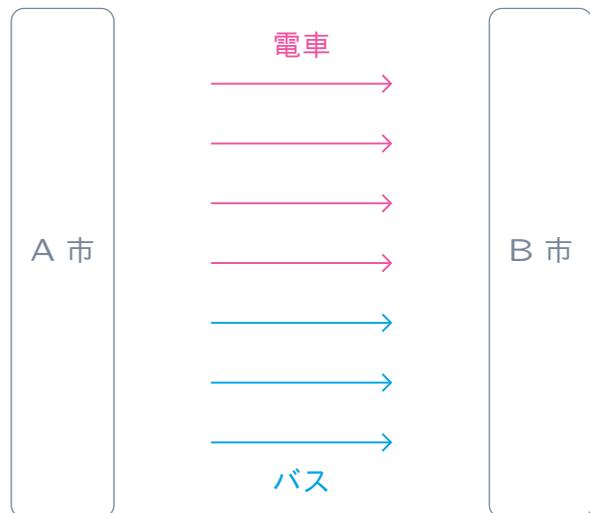
ここでは、和の法則と積の法則という最も基本的な法則について確認しておこう。

和の法則

たとえば、A 市から B 市まで行ける路線が、

- 電車で 4 路線
- バスで 3 路線

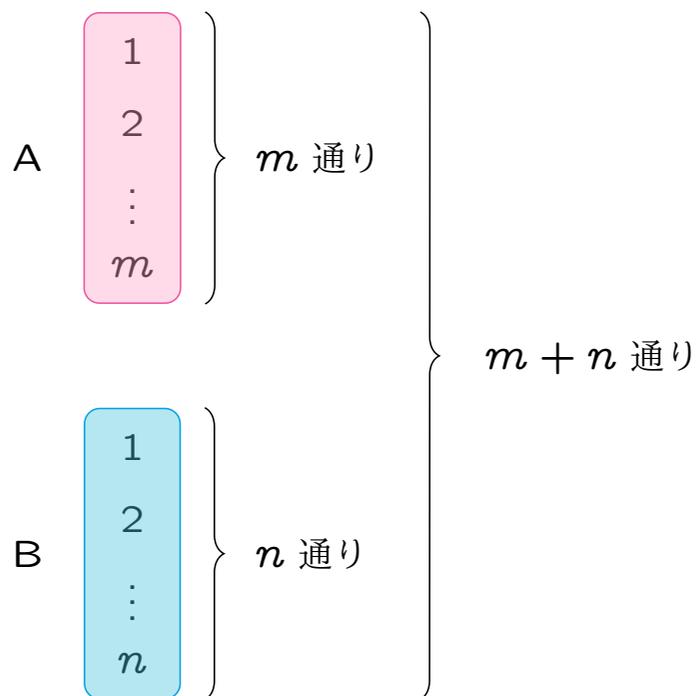
あるとする。



このとき、電車かバスの「どちらか」で A 市から B 市まで行くときには、 $4 + 3 = 7$ パターンの路線から選ぶことになる。

⚓ 和の法則 A と B は同時に起こらないとする。A の起こり方が m 通り、B の起こり方が n 通りあるとき、

A と B のどちらかが起こる場合は $m + n$ 通り



積の法則

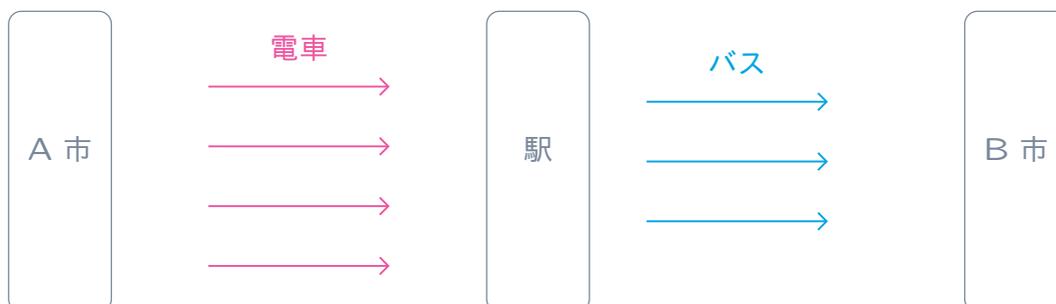
今度は、A 市から B 市へ、駅を経由して行く場合を考えてみる。

A 市から駅までは電車で、駅から B 市まではバスで行くとする。

つまり、電車とバスを「両方使って」移動することになる。

- A 市から駅までの電車は 4 路線
- 駅から B 市までのバスは 3 路線

どの路線の電車で行くかを決めたら、今度はどの路線のバスに乗るかを選ぶことになる。



4 通りの中からどの路線の電車を選んでも、次に乗るバスは 3 通りの中から選ぶ必要があるので、電車の路線 1 つにつき、次に乗るバスの路線は 3 パターン考えられる。

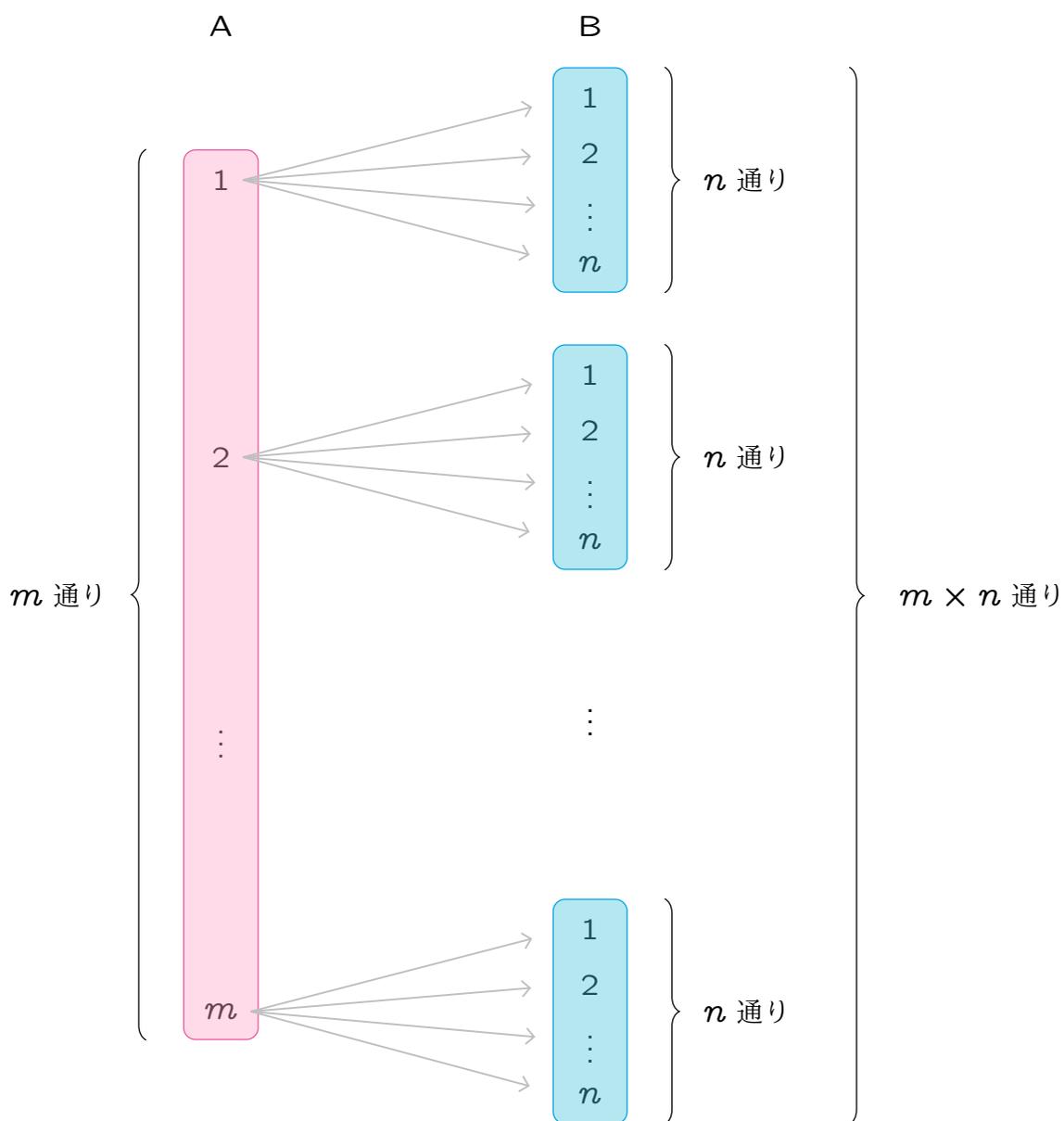
「電車 1 路線につきバス 3 路線」というパターンの数は、かけ算で表すことができそうだ。

電車とバスを乗り継ぐ場合の路線の選び方は、 $3 \times 4 = 12$ 通りになる。

 積の法則 A の起こり方が m 通りあり、その各々について B の起こり方が n 通り考えられるとき、

A と B がともに起こる場合は mn 通り

A と B が「ともに起こる」とは、A が起こった後に B が起こる場合を指す。



和事象の確率

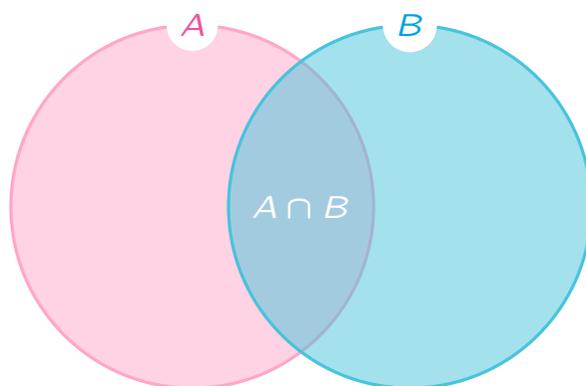
確率は論理の拡張であると捉えると、AND や OR といった論理演算を確率に当てはめて考えることができる。ここからは、複数の出来事（事象）が組み合わせられた場合の確率について考えていこう。

まずは OR「または」で組み合わせられた事象の確率を考えてみる。

A または B が起こる事象は、 $A \cup B$ と表すことができる。このような事象を和事象という。

A または B が起こる場合が $n(A \cup B)$ 通りあるとすると、その確率は、次の割合で表される。

$$P(A \cup B) = \frac{n(A \cup B)}{n(U)}$$



ここで、 $n(A \cup B)$ は、場合の数の和の法則より、 $n(A)$ と $n(B)$ の和で求まると考えられる。しかし、 A と B の重なっている部分は二重に数えてしまうので、 $n(A \cap B)$ を引く必要がある。

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

よって、和事象 $A \cup B$ の確率は、

$$\begin{aligned} P(A \cup B) &= \frac{n(A \cup B)}{n(U)} \\ &= \frac{n(A) + n(B) - n(A \cap B)}{n(U)} \\ &= \frac{n(A)}{n(U)} + \frac{n(B)}{n(U)} - \frac{n(A \cap B)}{n(U)} \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

として求められる。

 和事象の確率 和事象 $A \cup B$ の確率は、

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



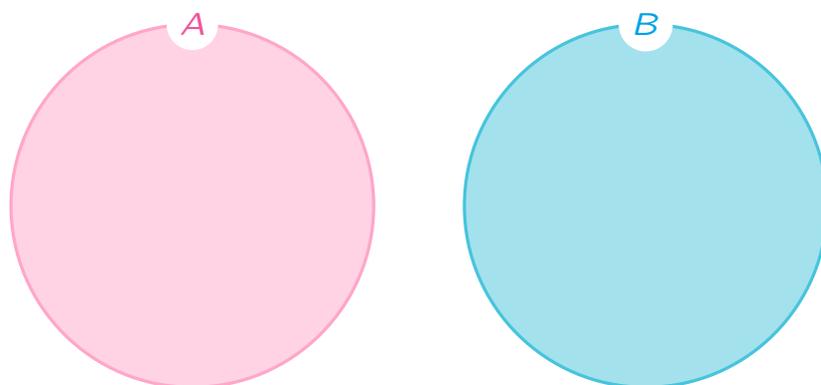
排反な事象と和事象の確率

2つの事象（出来事）が互いに排反であるとは、

一方の出来事が起こると、もう一方の出来事は起こりえない



という状況を指す。



和事象の確率において、事象 A と B が互いに排反であるなら、 $n(A \cap B) = 0$ となるので、

$$P(A \cup B) = P(A) + P(B)$$

が成り立つ。これを**確率の加法定理**という。

 確率の加法定理 互いに排反な事象 A, B の和事象 $A \cup B$ の確率は、

$$P(A \cup B) = P(A) + P(B)$$

条件つき確率

事象 A が起こったときに事象 B が起こる確率を **条件つき確率** という。

 **条件つき確率** 事象 A が起こったときに事象 B が起こる確率を $P(B|A)$ あるいは $P_A(B)$ と表し、これを A が起こったときの B が起こる **条件つき確率** という。

条件つき確率では、標本空間「全体」ではなく、その一部分である「 A が起きた場合」に限定して考える。

その中で B も起こる割合だから、「 A かつ B 」の確率を「 A 」の中での割合でみればよい。

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

 **条件つき確率** 事象 A が起こったときの事象 B が起こる条件つき確率は、

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

積事象の確率

ここで取り上げるのは、AND「かつ」で組み合わせられた事象の確率である。

A かつ B が起こる事象は、 $A \cap B$ と表すことができる。このような事象を **積事象** という。

A かつ B が起こる場合が $n(A \cap B)$ 通りあるとすると、その確率は、次のような割合で表される。

$$P(A \cap B) = \frac{n(A \cap B)}{n(U)}$$

一方、積事象の確率は、次のように分けて考えることもできる。

1. 全体のうち A が起こる (この確率は $P(A)$)
2. A が起こったとき、 B が起こる (この確率は $P(B|A)$)

全体のうち A が起こる場合の数を $n(A)$ 、 A が起こった場合のうち B が起こる場合の数を $n(B|A)$ とすると、**場合の数の積の法則**より、

$$n(A \cap B) = n(A) \cdot n(B|A)$$

よって、積事象 $A \cap B$ の確率は、

$$\begin{aligned} P(A \cap B) &= \frac{n(A \cap B)}{n(U)} \\ &= \frac{n(A) \cdot n(B|A)}{n(U)} \\ &= \frac{n(A)}{n(U)} \cdot \frac{n(B|A)}{n(U)} \\ &= P(A) \cdot P(B|A) \end{aligned}$$

という形で表すことができる。これを**確率の乗法定理**という。

 **確率の乗法定理** 積事象 $A \cap B$ の確率は、

$$P(A \cap B) = P(A) \cdot P(B|A)$$

独立な事象と積事象の確率

2つの事象（出来事）が**互いに独立**であるとは、

一方の出来事の結果が、もう一方の出来事の結果に影響を与えない

ということである。

このとき、 A が起きたかどうかは B の起きやすさに影響しないので、

$$P(B|A) = P(B)$$

が成り立つ。

よって、確率の乗法定理を次のように書き換えられる。

$$P(A \cap B) = P(A) \cdot P(B)$$

 独立な事象の確率 互いに独立な事象 A, B の積事象 $A \cap B$ の確率は、

$$P(A \cap B) = P(A) \cdot P(B)$$

第 4 章

確率分布

記述統計から推測統計へ

ここまで扱ってきた、代表値やばらつきの指標、度数分布などは、全データをもとに算出されるものだった。

すべてのデータをもとに、それらを整理することでデータ全体の性質を分析する方法論は、**記述統計**と呼ばれる。

しかし、調べたい対象の規模によっては、それらすべてからデータを集めるのは不可能な場合もある。

たとえば、日本人の平均身長を求めようとしても、日本人全員に身長を聞いて回ることは現実的ではない。

そこで、一部のデータだけを集めて、そこからデータ全体の性質を推測する方法論として、**推測統計**がある。

母集団からの無作為抽出

推測統計では、調べたいデータ全体を**母集団**といい、その一部分だけを使って推測を行う。

推測に使う、一部分のデータを**標本**という。

そして、母集団から標本を取り出すことを、**標本抽出**という。

たとえば、日本人の平均身長を推測するためのデータを集めたいとする。

友達に身長を聞いて回ってデータを集めるのも標本抽出といえるが、これでは同性の身長データが比較的多く集まってしまうなど、偏りが生じてしまう。

ある特定の対象に偏ることなくデータを集めるためには、ランダムに聞いて回る必要がある。

ランダムに標本を抽出することを、**無作為抽出**という。



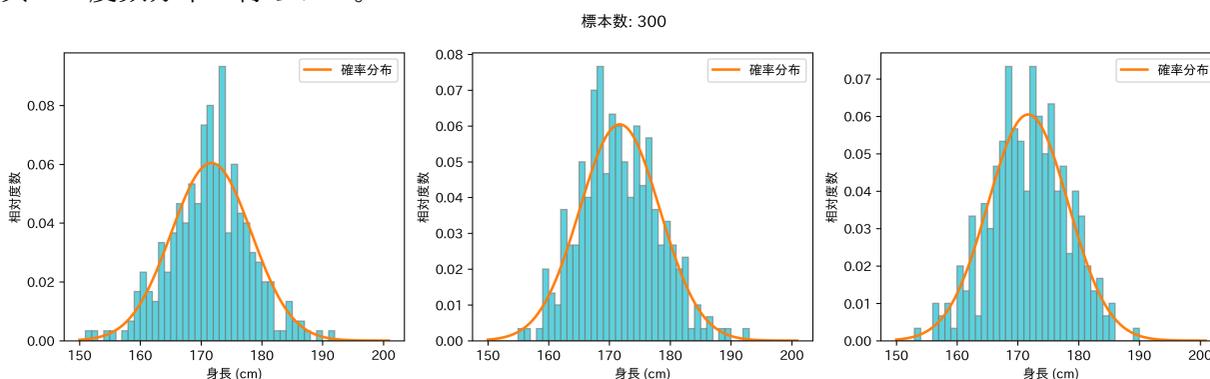
母集団の度数分布

無作為抽出はランダムなので、どんな標本が得られるかはわからない。

このランダム性は、**確率**を用いてデータを解釈するという考え方に結びついていく。

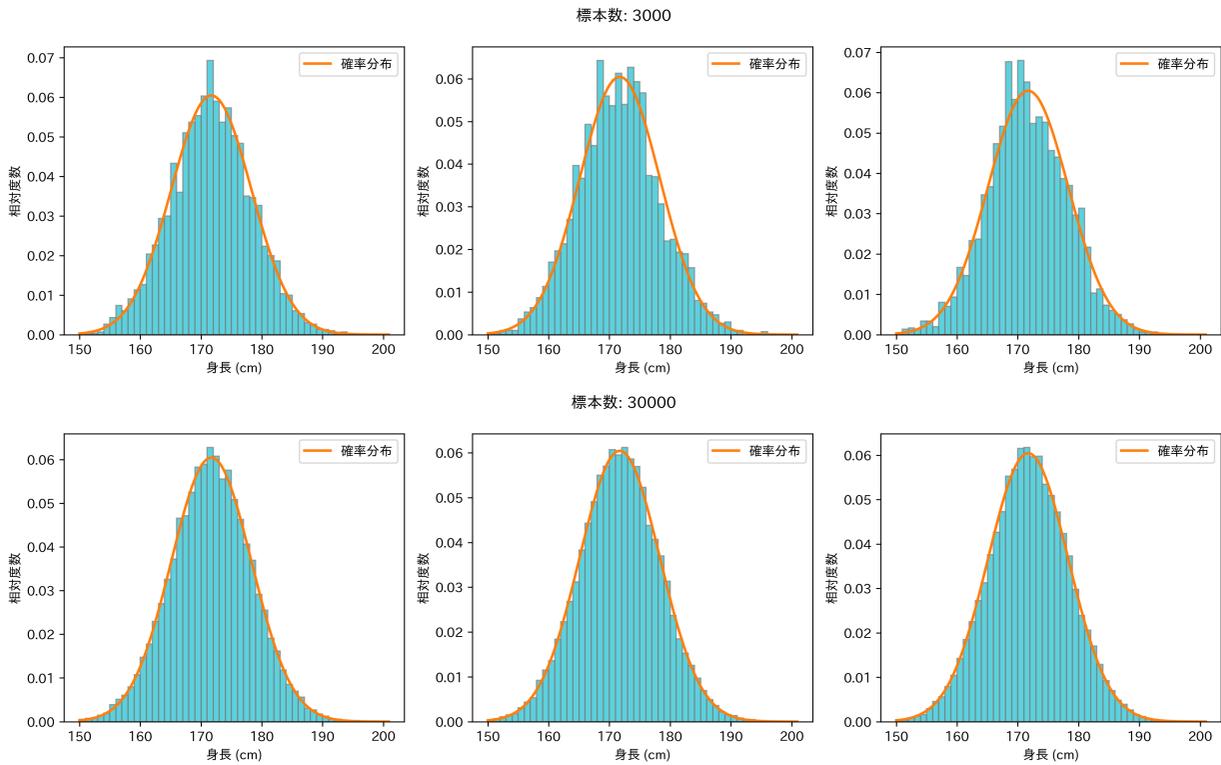
無作為抽出して得られた身長データを 1cm ごとの階級幅に区切って、(相対)度数分布として整理したとしよう。

無作為抽出は行うたびに結果が異なるため、抽出する標本数が少なければ、抽出を行うたびに毎回異なる度数分布が得られる。



しかし、この揺らぎは、標本数を大きくしていくと小さくなっていく。

すなわち、十分多くの標本を無作為抽出すれば、何度抽出を繰り返しても、得られる度数分布は一定の分布に近づく。



抽出する標本の数を大きくしていけばいくほど、抽出した「一部分」は「全体」に近くなっていくため、その度数分布は母集団の度数分布に近づいていくことになる。



相対度数と確率

ここで、相対度数は、全体において各階級の値が現れる割合を表していた。

この割合は、無作為抽出のもとでは「全体の中である階級値のデータが現れる確率」と捉えることができる。

たとえば、173cm 台という階級の相対度数が 0.25 だったとする。

このとき、抽出に偏りがなから、173cm 台という値が母集団全体の 25% を占めているといえるので、母集団全体から標本を取り出したときに 173cm 台という値が得られる確率も 25% となる。

無作為抽出では、相対度数は確率を表す



実際、相対度数 $\frac{f_i}{N}$ の合計は 1 となるので、確率の定義も満たしている。

$$\sum_{i=1}^N \frac{f_i}{N} = \frac{1}{N} \sum_{i=1}^N f_i = \frac{1}{N} \cdot N = 1$$

標本と確率変数

ここで、 X という値が得られる確率を関数に見立てて、次のように表そう。

$$P(X)$$

173cm 台という値が得られる確率は、次のように書ける。

$$P(X = 173\text{cm 台}) = 0.25$$

この関数の引数となる変数 X を、**確率変数**という。

確率変数は、「結果を言い当てることはできないが、とりうる値とその値が出る確率が決まっている」ものをいう。

 **確率変数** 取りうる値すべてに対して確率を考えることができるような変数

ここで、「173cm 台という値が得られた」ということは、「取り出した標本が 173cm 台だった」ということなので、 X は**標本**である。

無作為抽出では、標本は確率変数とみなせる



母集団の度数分布と確率分布

この $P(X)$ は、各値 X がどれくらいの確率で出るかを表す関数といえる。

X の取りうる値それぞれに対して確率を割り振り、すべての確率を合計すると 1 となる。

$P(X)$ は、この確率の割り振り方、すなわち全確率 1 が確率変数 X の取りうる値にどのように分布しているかを表しているため、 X の確率分布と呼ばれる。

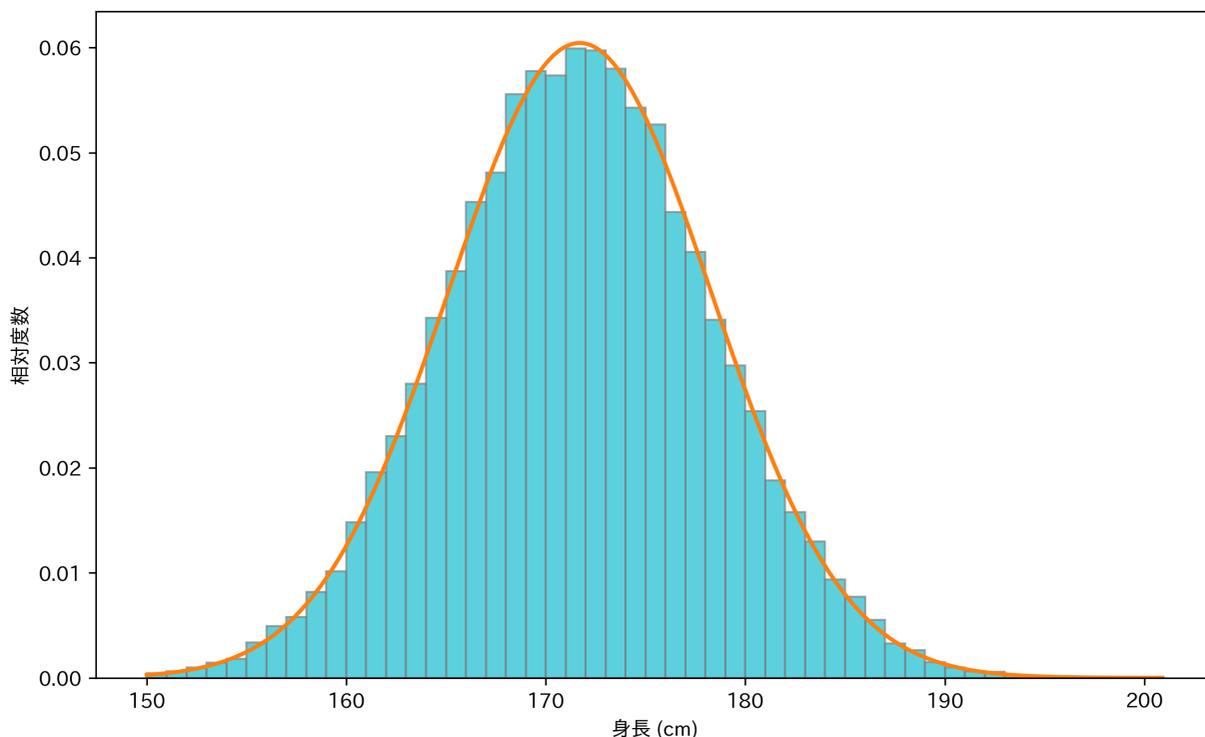
また、母集団のある階級の相対度数は、その母集団から無作為抽出された標本 X がその階級に属する確率 $P(X)$ を表していた。

無作為抽出では、母集団の度数分布は標本の確率分布となる



連続型確率分布

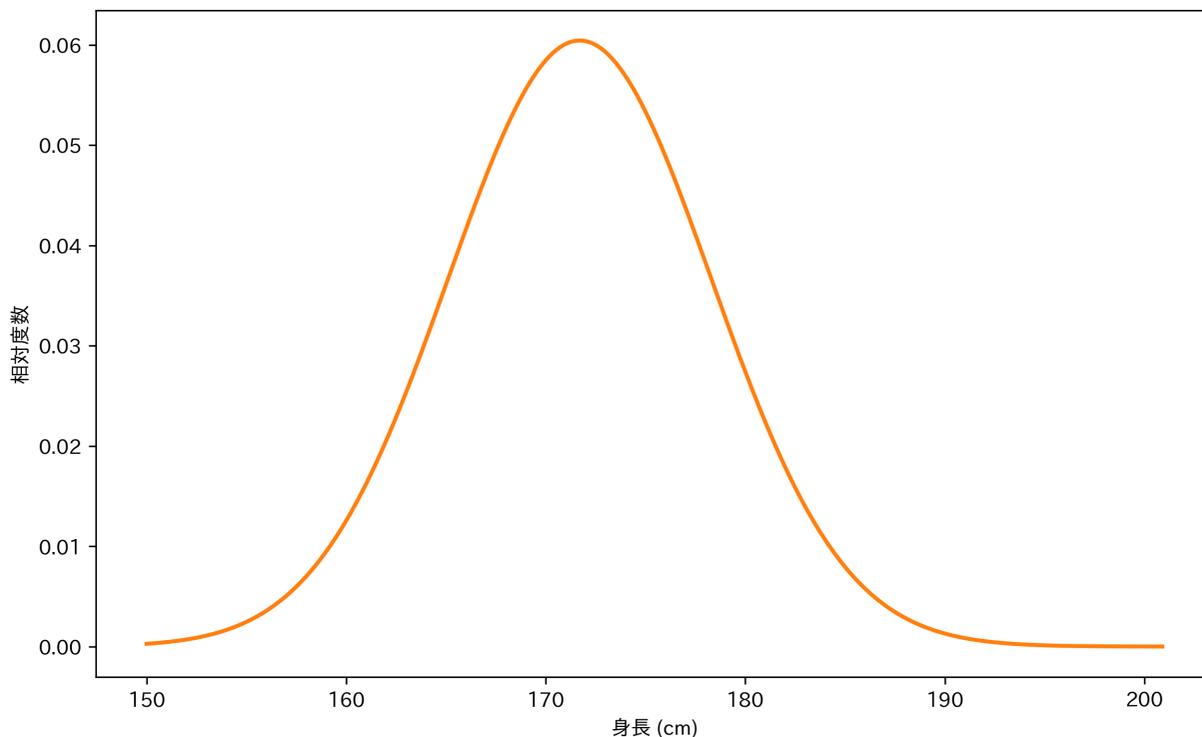
階級幅を横幅、相対度数を高さとする棒グラフ（ヒストグラム）を作成すると、これはそのまま標本 X の確率分布 $P(X)$ を表すグラフとなる。



しかし、たとえば 173cm 台を表す棒だけでは、173.1cm の人が何人いて、173.5cm の人が何人いたのかなど、その間の分布はわからない。

そこで、棒グラフの幅（階級幅）をどんどん細かくしていくことで、173.8...cm といった実際の身長に近い値（実数値）をとる確率の分布を考えることはできないだろうか？

階級幅を無限に細かく分割した場合、それぞれの棒は幅が限りなく 0 に近い「線」となる。この線の高さを結んでいくと、次のようなグラフが浮かび上がるだろう。



しかし、階級幅を小さくしていくと、各階級に含まれるデータの数（度数）はどんどん少なくなっていく。これは、相対度数（確率）の式 $\frac{f_i}{N}$ でいえば、度数 f_i がどんどん小さくなっていくことに相当する。

$$\lim_{f_i \rightarrow 0} \frac{f_i}{N} = 0$$

このように、連続的なデータ（実数値をとるデータ）では、完全に一致する値（特定の 1 点）が出る確率は 0 となってしまう。

だから「173cm 台」すなわち「173cm 以上 174cm 未満」といった区間に対して確率を考えていたのである。

確率変数 X が実数値をとるような確率分布は、**連続型確率分布**と呼ばれる。

連続型確率分布では、「ある区間に入る確率」を考えるしかない。

$P(X = 173\text{cm 台}) = 0.25$ という確率の式は、区間に入る確率として、次のように書き換え

られる。

$$P(173 \leq X < 174) = 0.25$$

確率密度関数

グラフを見ると、たとえば 172.0cm というデータが得られる確率（相対度数）と、172.5cm というデータが得られる確率は、ほぼ同じだといえる。

このように、区間幅 Δx が十分に小さいとき、 $X = x$ での確率と、 $X = x + \Delta x$ での確率はほぼ同じになる。

そこで、幅 Δx の棒を立てたとき、その棒（ Δx は微小なのでほとんど線に見える）の高さは、ほぼ 1 点 $X = x$ での確率と見なすことができる。

ほぼ 1 点 $X = x$ での確率を高さ $f(x)$ とみなせるなら、そこに区間幅 Δx をかけることで、区間に対する確率が求められる。

$$P(x \leq X < x + \Delta x) = f(x)\Delta x$$

幅 Δx の棒を、区間 $[a, b)$ を埋め尽くすように並べたとして。

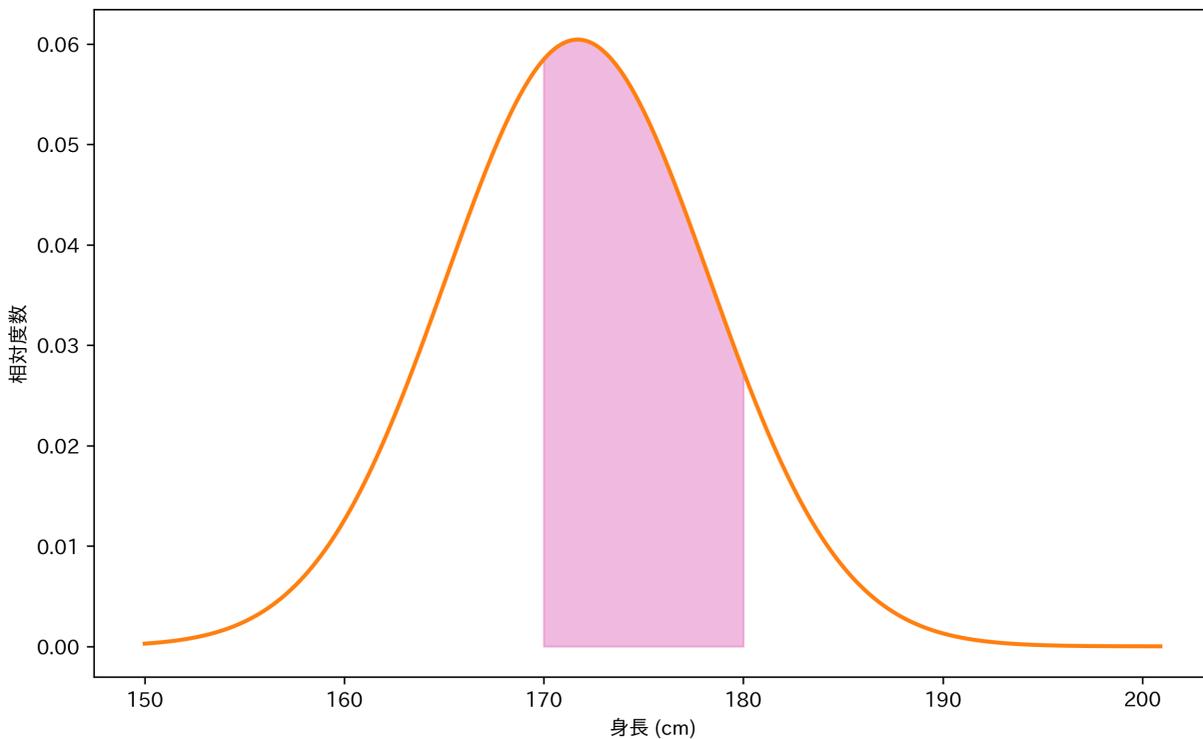
しかし、区間 $[a, b)$ 内で確率が変化しないとは限らないので、それぞれの棒の高さ $f(x)$ は X の値 x によって異なる。

そこで、「 x を変化させながら $f(x)\Delta x$ を足し合わせる」という演算が必要になる。

このような演算は、**定積分**として定義される。

$$P(a \leq X < b) = \int_a^b f(x)dx$$

たとえば、170cm 以上 180cm 未満の区間に対する確率 $P(170 \leq X < 180)$ は、次のピンクの領域の面積として求められる。



このように、連続型確率分布では、定積分（面積）によって区間に対する確率を定義する。

積分すると確率が求まるような関数 $f(x)$ を**確率密度関数**（PDF: probability density function）と呼ぶ。

確率の総和は 1 であるから、全区間に対する確率密度関数の定積分（全面積）は 1 となる必要がある。確率密度関数は、次の性質を満たすものとして定義される。

$$\int_{-\infty}^{\infty} f(x)dx = 1, \quad f(x) \geq 0$$

離散型確率分布

X が実数値をとる連続型確率分布に対して、 X が整数などの離散値をとる場合の確率分布を**離散型確率分布**と呼ぶ。

離散値とは、たとえばサイコロの目 1, 2, 3, 4, 5, 6 のように、値が飛び飛びで連続していないものをいう。

サイコロの目を確率変数 X とすると、 X が 1 の目を出す確率は、次のように表される。

$$P(X = 1) = \frac{1}{6}$$

確率質量関数

一般に、確率変数 X が値 x_i をとる確率が次のように表される時、

$$P(X = x_i) = f(x_i)$$

ここで現れた $f(x)$ は確率変数 X の確率質量関数 (PMF: probability mass function) と呼ばれる。