

第 160 講 正規分布

学習目標

この講では、正規分布について次のことができるようになることを目標とする。

- 正規分布 $N(\mu, \sigma^2)$ における平均 μ と標準偏差 σ の意味を説明できる。
- 確率変数を標準化して、標準正規分布を用いた確率計算へ変換できる。
- 標準正規分布表 $A(z) = P(0 \leq Z \leq z)$ を使い、対称性や補集合を利用して確率を求められる。
- 密度の高さと確率の違いを理解し、連続型確率変数の特徴を説明できる。

正規分布を考える理由

同じ種類のデータを集めると、平均付近に多くの値が集まり、平均から離れるほど少なくなる形が現れることがある。

ただし、すべてのデータが必ずこの形になるわけではない。正規分布は、ある現象を説明するための数学的なモデルとして「この確率変数は正規分布に従う」と仮定して使うものである。

正規分布を利用すると、正規分布に従うと仮定したモデルの分布は、平均と正の標準偏差によって定まり、そのばらつきを表現できる。

正規分布で重要なのは、実際の値そのものを見るのではなく、「平均から標準偏差何個分離れているか」に変換することである。

正規分布の定義

正規分布とは、平均 μ 、標準偏差 σ の連続型確率分布であり、次の式で密度関数が表される分布である。

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

ここで、 x と μ は実数、 $\sigma > 0$ であり、 $e \approx 2.71828$ は自然対数の底である。

この式は正規分布を特定するために示したものであり、この講では密度関数の微積分やモーメントの導出は扱わない。

正規分布では、平均が μ 、分散が σ^2 、標準偏差が σ であり、密度曲線全体の面積は 1 になる。

また、平均 μ を中心として左右対称である。

特に、 $\mu = 0$ 、 $\sigma = 1$ のときの正規分布を標準正規分布という。

標準正規分布は、 $Z \sim N(0, 1)$ と表し、標準正規分布表を用いた確率計算の基準になる。

正規分布の記法

一般に、平均 μ 、分散 σ^2 の正規分布を、

$X \sim N(\mu, \sigma^2)$ と表す。

このとき、第 2 引数 σ^2 は分散を表し、 $\sigma > 0$ が標準偏差である。

正規分布で覚える性質

正規分布では、次の性質を利用する。

- 平均 μ を中心として左右対称である。

- 密度曲線全体の面積は 1 である。
- 平均が同じ正規分布では、標準偏差 σ が大きいほど横に広がり、最大の密度は低くなる。

密度の高さは、その点そのものの確率を表しているわけではない。

ある短い幅の区間について考えると、同じ幅の区間では密度が高い場所ほど面積が大きくなりやすく、確率も大きくなりやすい。この考え方は、区間の確率を考える手がかりである。

正規密度はどの有限実数 x においても正である。

一方で、 $P(X = c) = 0$ は、 c が実数の 1 点だけでは曲線の下側の面積を持たないことを表している。

したがって、1 点の確率が 0 であるという情報だけから、その値が論理的に不可能であるとは結論できない。

区間の端点を含めるかどうかによって確率が変わらないのも、この性質による。

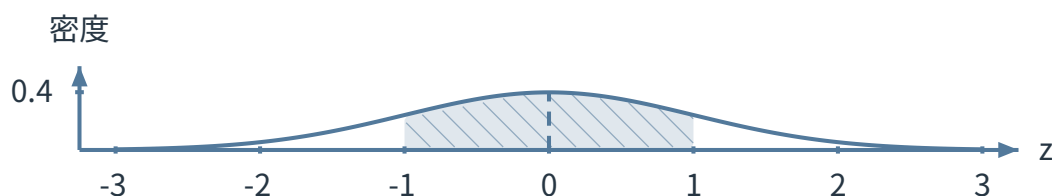


図1 標準正規分布の密度を示す。平均 0 を中心に左右対称で、斜線部は $-1 \leq Z \leq 1$ の確率を表す。描画した $-3 \leq z \leq 3$ は全実数上の曲線の一部であり、両端は定義域の端点ではない。

標準化という考え方

正規分布では、値そのものではなく、平均から標準偏差何個分離れているかを見る。

例えば、 $X \sim N(50, 5^2)$ と $Y \sim N(100, 10^2)$ を考える。

X で平均より 10 大きい値は 60 であり、

$$\frac{60 - 50}{5} = 2$$

なので、平均から標準偏差 2 個分上にある。

一方、 Y で平均より 10 大きい値は 110 であり、

$$\frac{110 - 100}{10} = 1$$

なので、平均から標準偏差 1 個分上にある。

このように、同じ差 10 でも標準偏差によって意味が変わる。

そこで、確率変数 X を、

$$Z = \frac{X - \mu}{\sigma}$$

によって変換する。

ここで、 $X - \mu$ は平均からの符号付き偏差を表す。

- $Z > 0$ のとき、値は平均より右側にある。
- $Z < 0$ のとき、値は平均より左側にある。
- $|Z|$ は平均から標準偏差何個分離れているかを表す。

$\sigma > 0$ なので、標準化によって不等号の向きは変化しない。

正規分布に従う確率変数では、この標準化によって、 $X \sim N(\mu, \sigma^2)$ から、 $Z \sim N(0, 1)$ へ変換できる。

標準正規分布表

標準正規分布 $Z \sim N(0, 1)$ について、この教材では次の表を利用する。

表の値は、 $A(z) = P(0 \leq Z \leq z)$ を表す。

ただし、表は $z \geq 0$ の範囲だけを扱う。

z	$A(z)$
0	0
0.5	0.1915
1	0.3413
1.5	0.4332
2	0.4772

$A(0) = 0$ は厳密な値である。

一方、 $A(0.5)$ 、 $A(1)$ 、 $A(1.5)$ 、 $A(2)$ について、表に載せたこれらの A の数値が、小数第 4 位に丸めた近似値である。

したがって、表から計算した確率は $\approx$ を用いて表し、厳密な値とは区別する。

表の使い方

標準正規分布表は、

$A(z) = P(0 \leq Z \leq z)$ であり、 $z \geq 0$ の値を利用する。

$0 \leq u \leq v$ のとき、左右対称性と面積の考え方から、 $P(-u \leq Z \leq u) = 2A(u)$ となる。

また、右側の裾の確率は、

$$P(Z \geq v) = \frac{1}{2} - A(v)$$

で求められる。

さらに、同じ右側の範囲の差を利用すると、 $P(u \leq Z \leq v) = A(v) - A(u)$ となる。

区間の端点を含めるかどうかは、連続型確率変数では 1 点の確率が 0 であるため、確率計算の結果に影響しない。

正規分布で注意すること

正規分布について、次の誤りに注意する。

- すべてのデータが正規分布になるわけではない。
- 正規分布を使う場合は、対象の確率変数が正規分布に従うと仮定している。
- 密度の高さと確率を混同しない。
- $N(\mu, \sigma^2)$ の第 2 引数は標準偏差ではなく分散である。
- 標準化では平均を引いてから標準偏差で割る。

例題 1

標準正規分布 $Z \sim N(0, 1)$ について、次の確率を求めよ。

(1) $P(-1 \leq Z \leq 1)$

(2) $P(Z \geq 1)$

例題 1 の解答・解説

(1) 標準正規分布は 0 を中心に左右対称である。

表より、 $A(1) = P(0 \leq Z \leq 1) \approx 0.3413$ である。

したがって、

$$\begin{aligned} P(-1 \leq Z \leq 1) &= 2P(0 \leq Z \leq 1) \\ &\approx 2 \times 0.3413 \\ &\approx 0.6826 \end{aligned}$$

よって、 $P(-1 \leq Z \leq 1) \approx 0.6826$ である。

(2) $Z \geq 1$ は、右側の端の部分である。

$Z \geq 0$ の確率は左右対称性より、

$$P(Z \geq 0) = \frac{1}{2}$$

である。

したがって、

$$\begin{aligned} P(Z \geq 1) &= P(Z \geq 0) - P(0 \leq Z \leq 1) \\ &\approx 0.5 - 0.3413 \\ &\approx 0.1587 \end{aligned}$$

よって、 $P(Z \geq 1) \approx 0.1587$ である。

判断のポイントは、表が 0 から z までの面積を示す表であるため、必要な範囲へ分解することである。

例題 2

確率変数 X が $X \sim N(50, 100)$ に従うとする。

(1) 平均と標準偏差を求めよ。

(2) $P(40 \leq X \leq 65)$ を求めよ。

例題 2 の解答・解説

(1) 正規分布の表し方

$X \sim N(\mu, \sigma^2)$ では、第 2 引数は分散である。

したがって、 $\mu = 50$ であり、 $\sigma^2 = 100$ である。

標準偏差は分散の平方根なので、 $\sigma = \sqrt{100} = 10$ である。

よって、平均は 50、標準偏差は 10 である。

(2) X を標準化する。

$$Z = \frac{X - 50}{10}$$

である。
範囲を変換すると、

$$\frac{40 - 50}{10} \leq Z \leq \frac{65 - 50}{10}$$

より、 $-1 \leq Z \leq 1.5$ となる。
したがって、 $P(40 \leq X \leq 65) = P(-1 \leq Z \leq 1.5)$ である。
左右に分けて考える。

$$P(-1 \leq Z \leq 1.5) = P(-1 \leq Z \leq 0) + P(0 \leq Z \leq 1.5)$$

対称性より、 $P(-1 \leq Z \leq 0) = A(1) \approx 0.3413$

また、 $A(1.5) \approx 0.4332$ である。

したがって、

$$\begin{aligned} P(-1 \leq Z \leq 1.5) &\approx 0.3413 + 0.4332 \\ &\approx 0.7745 \end{aligned}$$

よって、 $P(40 \leq X \leq 65) \approx 0.7745$ である。

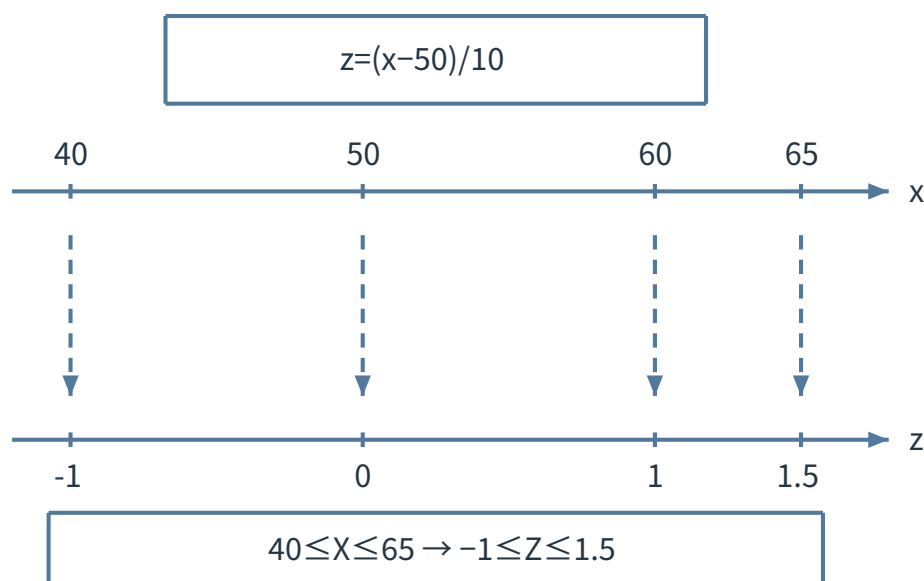


図2 Xは平均50、分散100、標準偏差10の正規分布に従う。同じ列は $z=(x-50)/10$ で対応し、 $40 \leq X \leq 65$ は $-1 \leq Z \leq 1.5$ に移る。上段と下段の単位は異なり、それぞれの軸内の間隔は数値の差に比例する。

例題3

確率変数 X が $X \sim N(100, 16)$ に従うとする。

- (1) $P(X \geq 106)$ を求めよ。
- (2) $P(X = 106)$ を求めよ。
- (3) 密度の値と確率の違いを説明せよ。

例題 3 の解答・解説

まず、

$\sigma^2 = 16$ なので、 $\sigma = 4$ である。

(1) $X = 106$ を標準化する。

$$Z = \frac{X - 100}{4}$$

より、

$$\frac{106 - 100}{4} = 1.5$$

である。

したがって、 $P(X \geq 106) = P(Z \geq 1.5)$ となる。

表より、 $A(1.5) \approx 0.4332$ なので、

$$\begin{aligned} P(Z \geq 1.5) &= \frac{1}{2} - A(1.5) \\ &\approx 0.5 - 0.4332 \\ &\approx 0.0668 \end{aligned}$$

よって、 $P(X \geq 106) \approx 0.0668$ である。

(2) 連続型確率変数では、特定の 1 点になる確率は 0 である。

したがって、 $P(X = 106) = 0$ である。

(3) 正規分布の曲線の高さは密度を表している。

密度が高い場所は、その近くに値が現れやすいことを意味する。しかし、密度の高さそのものが確率ではない。

確率は、ある区間に対応する曲線の下側の面積で表される。

したがって、点の高さを見るのではなく、区間の面積を見ることが重要である。

演習問題

練習問題 1

標準正規分布 $Z \sim N(0, 1)$ について、次の確率を求めよ。

(1) $P(0 \leq Z \leq 0.5)$

(2) $P(-0.5 \leq Z \leq 0.5)$

練習問題 1 の解答・解説

(1) 標準正規分布表は、

$A(z) = P(0 \leq Z \leq z)$ を表す。

したがって、 $z = 0.5$ の値をそのまま読む。

表より、 $A(0.5) \approx 0.1915$ であるから、 $P(0 \leq Z \leq 0.5) \approx 0.1915$ となる。

(2) 標準正規分布は 0 を中心に左右対称である。

したがって、左側の -0.5 から 0 までの面積は、0 から 0.5 までの面積と等しい。

よって、

$$\begin{aligned} P(-0.5 \leq Z \leq 0.5) &= 2P(0 \leq Z \leq 0.5) \\ &\approx 2 \times 0.1915 \\ &\approx 0.3830 \end{aligned}$$

である。

表の値は近似値なので、結果も $\approx$ で表す。

練習問題 2

標準正規分布 $Z \sim N(0, 1)$ について、次の確率を求めよ。

(1) $P(Z \leq -1.5)$

(2) $P(1 \leq Z \leq 2)$

練習問題 2 の解答・解説

(1) 左右対称性を利用する。

$Z \leq -1.5$ の面積は、 $Z \geq 1.5$ の面積と等しい。

表より、 $A(1.5) \approx 0.4332$ である。

曲線下全体の面積は 1 であり、標準正規分布は 0 を中心に左右対称なので、右半分の面積は、

$$P(Z \geq 0) = \frac{1}{2}$$

となる。

したがって、

$$\begin{aligned} P(Z \geq 1.5) &= \frac{1}{2} - P(0 \leq Z \leq 1.5) \\ &\approx 0.5 - 0.4332 \\ &\approx 0.0668 \end{aligned}$$

となる。

したがって、 $P(Z \leq -1.5) \approx 0.0668$ である。

(2) $1 \leq Z \leq 2$ は、0 から 2 までの面積から 0 から 1 までの面積を引く。

表より、 $A(2) \approx 0.4772$

$A(1) \approx 0.3413$ である。

したがって、

$$\begin{aligned} P(1 \leq Z \leq 2) &= P(0 \leq Z \leq 2) - P(0 \leq Z \leq 1) \\ &\approx 0.4772 - 0.3413 \\ &\approx 0.1359 \end{aligned}$$

である。

練習問題 3

確率変数 X が $X \sim N(20, 9)$ に従うとする。

(1) 平均と標準偏差を求めよ。

(2) $P(17 \leq X \leq 23)$ を求めよ。

練習問題 3 の解答・解説

(1) 正規分布

$X \sim N(\mu, \sigma^2)$ では、第 2 引数は分散である。

したがって、 $\mu = 20$

$\sigma^2 = 9$ である。

標準偏差は、 $\sigma = \sqrt{9} = 3$ となる。

よって、平均は 20、標準偏差は 3 である。

(2) 標準化する。

$$Z = \frac{X - 20}{3}$$

である。

範囲を変換すると、

$$\frac{17 - 20}{3} \leq Z \leq \frac{23 - 20}{3}$$

より、 $-1 \leq Z \leq 1$ となる。

したがって、 $P(17 \leq X \leq 23) = P(-1 \leq Z \leq 1)$ である。

表より、 $A(1) \approx 0.3413$ なので、

$$\begin{aligned} P(-1 \leq Z \leq 1) &= 2A(1) \\ &\approx 2 \times 0.3413 \\ &\approx 0.6826 \end{aligned}$$

よって、 $P(17 \leq X \leq 23) \approx 0.6826$ である。

練習問題 4

確率変数 X が $X \sim N(60, 25)$ に従うとする。

(1) $P(X \geq 70)$

(2) $P(X < 60)$ を求めよ。

練習問題 4 の解答・解説

分散が 25 なので、

$\sigma = \sqrt{25} = 5$ である。

(1) $X = 70$ を標準化する。

$$Z = \frac{X - 60}{5}$$

より、

$$\frac{70 - 60}{5} = 2$$

である。

したがって、 $P(X \geq 70) = P(Z \geq 2)$ となる。

表より、 $A(2) \approx 0.4772$ なので、

$$\begin{aligned}P(Z \geq 2) &= \frac{1}{2} - A(2) \\&\approx 0.5 - 0.4772 \\&\approx 0.0228\end{aligned}$$

よって、 $P(X \geq 70) \approx 0.0228$ である。

(2) $X < 60$ は平均より左側の範囲である。

連続型確率変数では、 $P(X < 60) = P(X \leq 60)$ である。

$X = 60$ という境界値を標準化すると、

$$Z = \frac{60 - 60}{5} = 0$$

となる。

これは確率変数 Z 全体が 0 になるという意味ではなく、境界値 $X = 60$ が標準化後の値 $z = 0$ に対応するという意味である。

また、事象 $X < 60$ は、標準化によって $Z < 0$ に対応する。

正規分布は左右対称なので、

$$P(Z \leq 0) = \frac{1}{2}$$

である。

したがって、

$$P(X < 60) = \frac{1}{2}$$

である。

練習問題 5

確率変数 X が $X \sim N(10, 4)$ に従うとする。

$P(7 \leq X \leq 9)$ を求めよ。

練習問題 5 の解答・解説

分散が 4 なので、
 $\sigma = 2$ である。
標準化すると、

$$Z = \frac{X - 10}{2}$$

となる。
範囲を変換する。

$$\frac{7 - 10}{2} \leq Z \leq \frac{9 - 10}{2}$$

より、 $-1.5 \leq Z \leq -0.5$ となる。
左右対称性を使って、 $P(-1.5 \leq Z \leq -0.5) = P(0.5 \leq Z \leq 1.5)$
と考える。
したがって、

$$\begin{aligned} P(0.5 \leq Z \leq 1.5) &= P(0 \leq Z \leq 1.5) - P(0 \leq Z \leq 0.5) \\ &\approx 0.4332 - 0.1915 \\ &\approx 0.2417 \end{aligned}$$

である。
よって、 $P(7 \leq X \leq 9) \approx 0.2417$ となる。
両端が平均より左側にあっても、左右対称性によって右側の同じ距離の範囲へ移して考えられる。

練習問題 6

確率変数 X が $X \sim N(30, 16)$ に従うとする。
ある人が、

$$Z = \frac{X - 30}{16}$$

として標準化した。
この誤りを説明し、正しい標準化を用いて
 $P(26 \leq X \leq 34)$ を求めよ。

練習問題 6 の解答・解説

分散は、

$\sigma^2 = 16$ である。

したがって、標準偏差は、 $\sigma = 4$ である。

標準化では、平均を引いた後、標準偏差で割る。

誤った式では 16 で割っているが、16 は分散であり標準偏差ではない。

正しい標準化は、

$$Z = \frac{X - 30}{4}$$

である。

次に範囲を変換する。

$$\frac{26 - 30}{4} \leq Z \leq \frac{34 - 30}{4}$$

より、 $-1 \leq Z \leq 1$ となる。

したがって、 $P(26 \leq X \leq 34) = P(-1 \leq Z \leq 1)$ である。

表より、 $A(1) \approx 0.3413$ だから、

$$\begin{aligned} P(-1 \leq Z \leq 1) &\approx 2 \times 0.3413 \\ &\approx 0.6826 \end{aligned}$$

よって、 $P(26 \leq X \leq 34) \approx 0.6826$ である。

練習問題 7

標準正規分布 $Z \sim N(0, 1)$ について、次の確率を求めよ。

(1) $P(Z = 0)$

(2) $P(Z \leq 0)$

(3) $P(Z > 0)$

また、 $z = 0$ において密度が最大になるのに、1 点の確率が 0 になる理由を説明せよ。

練習問題 7 の解答・解説

(1) 連続型確率変数では、特定の 1 点に対応する面積はない。

したがって、 $P(Z = 0) = 0$ である。

(2) 標準正規分布は 0 を中心に左右対称である。

全体の面積は 1 なので、

$$P(Z \leq 0) = \frac{1}{2}$$

である。

(3) 同様に、

$$P(Z > 0) = \frac{1}{2}$$

である。

密度は $z = 0$ において最大になるが、これは曲線の高さを表しているだけであり、1 点に対応する確率を表しているわけではない。

確率は曲線の下側の面積で決まり、1 点だけでは横幅がないため面積を持たない。

したがって、密度が最大の点であっても、 $P(Z = 0) = 0$ となる。

練習問題 8

次の 2 つの主張について、正しいか誤りかを判定し、理由を説明せよ。

(1) 「平均と標準偏差が与えられただけで、その確率分布が正規分布だといえる。」

(2) 「平均で引き、標準偏差で割ると、どんな分布でも標準正規分布になる。」

練習問題 8 の解答・解説

(1) 誤りである。

平均と標準偏差が与えられても、分布の形は決まらない。

例えば、長さが正の有限区間 $[a, b]$ ($a < b$) 上で一様に値をとる非退化な有限区間一様分布を考える。

この分布は有限の平均と有限かつ正の標準偏差を持つが、標準化後も有限区間外の確率は 0 である。

一方、非退化な正規分布では全実数上で密度が正であるため、この分布は正規分布ではない。

したがって、平均と標準偏差だけでは正規分布であるとは判断できない。

正規分布を用いる場合は、「この確率変数は正規分布に従う」と仮定する必要がある。

(2) 誤りである。

平均 μ が有限で、標準偏差 σ が有限かつ正である場合、標準化は、

$$Z = \frac{X - \mu}{\sigma}$$

という操作である。

標準化によって平均は 0、標準偏差は 1 になる。

しかし、標準化は分布の位置と尺度を変換する操作であり、分布の形を必ず正規分布に変える操作ではない。

正規分布に従う場合だけ、 $X \sim N(\mu, \sigma^2)$ から、 $Z \sim N(0, 1)$ となる。

したがって、「正規分布であるという前提」と「標準化という操作」は別のものである。