正規分布の入口

1　導入

この講義で最重要なのは、**正規分布は「平均のまわりに値が集まり、左右対称に広がる」分布の代表例だと捉えること**です。

統計で平均や標準偏差を学んだあと、つぎに大切になるのは「実際のデータがどんな形に散らばるか」です。そこで中心が一番出やすく、離れるほど出にくくなる分布として正規分布を見ます。

2　用語と定義

正規分布 Normal distribution とは、平均 μ、標準偏差 σ によって形が決まる、左右対称な連続分布です。

標準化 Standardization とは、

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

によって、平均 0、標準偏差 1 の形へ直すことです。

3　方針

まず正規分布を「平均のまわりに山がある分布」として直感的に捉えます。そのあと、平均と標準偏差が何を決めるかを整理し、最後に標準化の意味を押さえます。

→ 講義 **確率分布の基本** lecture math statistics
https://study.bem130.com/lecture/math/statistics/確率分布の基本-講義/

4　直感的な説明

身長、試験点、測定誤差などは、極端に大きい値や小さい値より、真ん中に近い値が出やすいことが多いです。この「中央が厚く、端へいくほど薄くなる」形を表す代表的な分布が正規分布です。

平均は山の中心を、標準偏差は山の広がりを決めます。つまり正規分布では、どこが中心かと、どれくらい散らばるかが分かれば、全体像のかなりの部分が見えてきます。

5　厳密な説明

5.1　1. 密度関数

正規分布の密度関数は

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

です。
ここで μ が平均、σ が標準偏差です。
この係数がなぜ

$$\frac{1}{\sqrt{2\pi}\sigma}$$

になるかは、全体の面積が 1 になるように決めているからです。つまり

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

となるように前の定数を選ばなければなりません。

5.2 1.5 ガウス積分

標準正規分布の核になる

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx$$

を直接積分するのは難しいですが、2 乗して

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy$$

と見ると平面の積分になります。ここで極座標 $x = r\cos\theta, y = r\sin\theta$ を使うと

$$I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\,dr\,d\theta$$

です。さらに $u = r^2/2$ と置換すると

$$\int_0^{\infty} e^{-r^2/2} r\,dr = \int_0^{\infty} e^{-u}\,du = 1$$

だから

$$I^2 = 2\pi$$

となり、

$$I = \sqrt{2\pi}$$

を得ます。これがガウス積分です。

したがって

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = 1$$

となるので、標準正規分布の密度関数は

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

でなければなりません。さらに $x = \mu + \sigma z$ と変数変換すると、一般の正規分布で

$$\frac{1}{\sqrt{2\pi}\sigma}$$

が現れる理由も分かります。

5.3 2. 左右対称

式の中に $(x-\mu)^2$ が入っているので、$x=\mu$ を中心として左右対称です。したがって平均の左右で同じだけ離れた値は、同じ出やすさを持ちます。

5.4 3. 標準化

平均 μ、標準偏差 σ の分布を

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

で変形すると、平均 0、標準偏差 1 の標準正規分布になります。

5.5 4. 統計で頻出する理由

細かい誤差や独立な要因がたくさん重なると、全体としては正規分布に近い形が現れやすいことが多いです。このため、統計や測定の話題で頻繁に出てきます。

6 見分け方

- 平均のまわりに左右対称な山形の分布が描かれていたら、正規分布を疑います。
- 標準化という言葉が出たら、$Z=\frac{X-\mu}{\sigma}$ を思い出します。
- 点数や偏差値の話題では、平均と標準偏差が主役です。

7 最終形

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

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

8 一言でいうと

- 正規分布は、平均のまわりに左右対称に集まる分布の基本形です。

9 関連リンク

→ 講義 確率分布の基本 lecture math statistics
https://study.bem130.com/lecture/math/statistics/確率分布の基本-講義/

→ 講義 **統計の基本** lecture math statistics
https://study.bem130.com/lecture/math/statistics/統計の基本-講義/