

P.74~P.77

基本事項

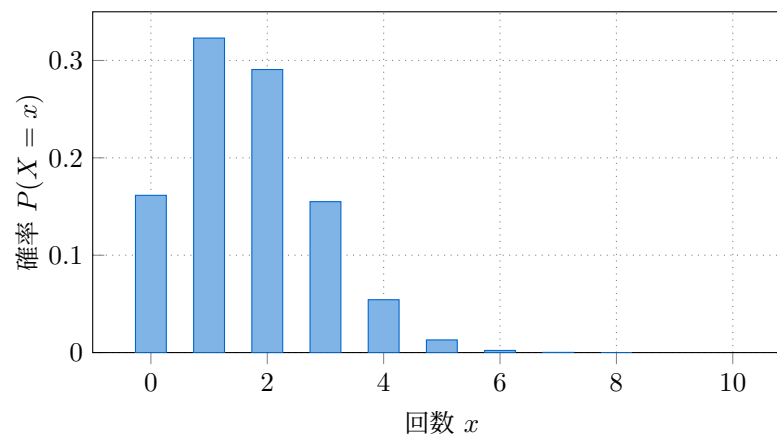
- (復習)「離散型」の確率変数  $X$  とは、「 $X$  の値が定まれば確率が決まる」ような  $X$  のこと.
  - 例 1 : コイン投げ「表なら  $X = 1$ 」, 「裏なら  $X = 0$ 」 $\rightarrow P(X = 1) = 1/2, P(X = 0) = 1/2$ .
  - 例 2 : サイコロ投げ「 $i$  の目が出たら  $X = i (i = 1, 2, \dots, 6)$ 」 $\rightarrow P(X = i) = 1/6$
- 「連続型」の確率変数  $X$  とは、「 $X$  の範囲が定まれば確率が決まる」ような  $X$  のこと.

説明

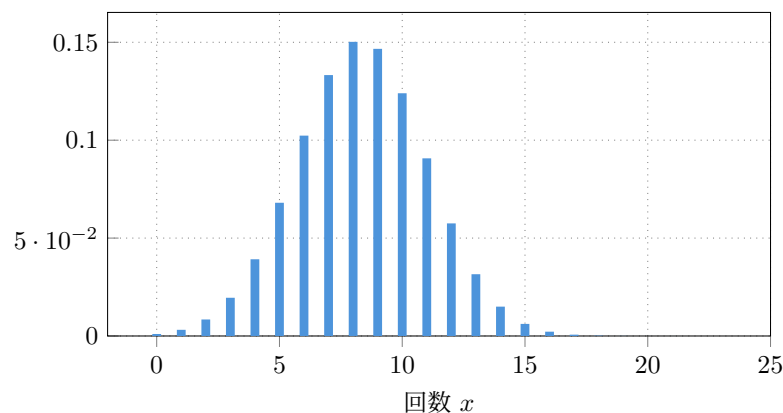
例 :  $X =$  身長

- 身長 170cm の人 ( $X = 170$ ) は【誰もいない / 何人か存在する】 $\rightarrow P(X = 170) = 0$ .
  - 身長 170cm 以上 171cm 未満の人 ( $170 \leq X < 171$ ) は【誰もいない / 何人か存在する】 $\rightarrow P(170 \leq X < 171) > 0$
  - 1 点「 $X = 170$ 」では確率が測れないから、「 $170 \leq X < 171$ 」のような範囲で確率を定める.
- 試行回数  $n$  を十分大きくすると, 右のヒストグラムのように, 形は徐々に滑らかになり, ある曲線  $y = f(x)$  が現れる. この曲線  $y = f(x)$  を分布曲線といい, 関数  $f(x)$  を確率密度関数という.

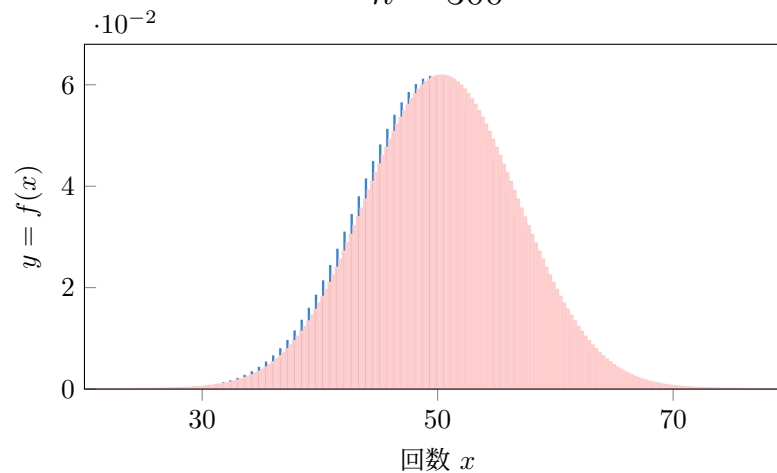
$n = 10$



$n = 50$



$n = 300$



全体課題 pre/post

ある工場で, ペットボトル飲料 (内容量  $X = 500\text{ml}$ ) を生産している. 機械の特性上, 内容量  $X$  にはわずかなばらつきが生じる. 過去のデータから, この内容量  $X$  は, 平均  $m = 500\text{ml}$ , 標準偏差  $\sigma = 2\text{ml}$  の「正規分布」という分布に従うことがわかっている. この工場では, 内容量が 498ml 未満, または 502ml 超過 の製品を「不良品」として出荷できない.

この工場で不良品が発生する確率  $P(X < 498 \text{ または } X > 502)$  を求めるための「考え方」を説明せよ (確率を求めるのは次回以降).

解答

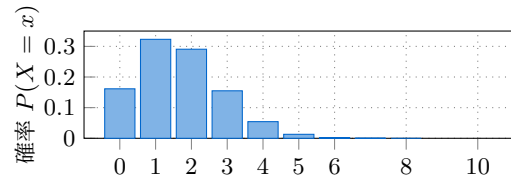
♠ エキスパート A 「面積と確率」

目標 A

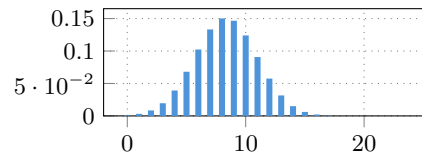
確率密度関数  $f(x)$  が確率の詰まり具合を表していることを理解し、確率が面積であることを説明できる。

例

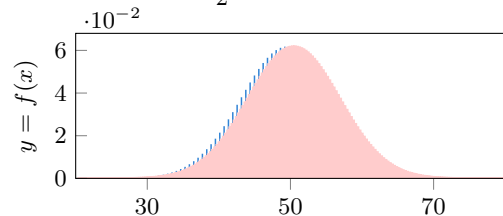
- $n$  回サイコロを振り、1 の目が出た回数を  $X$  とすると、 $X$  は、二項分布  $B\left(n, \frac{1}{6}\right)$  に従う。
  - $n = 10$  のとき、期待値は  $E(X) = \underline{\hspace{1cm}}$  だから、大体 1,2 回前後の確率が大きくなっている。さらに、各棒の横幅は 1、縦幅はそれぞれの確率を表しているため、棒の面積の合計は  $\underline{\hspace{1cm}}$ 。



- $n = 50$  のとき、期待値は  $E(X) = \underline{\hspace{1cm}}$  だから、大体 8,9 回前後の確率が大きくなっている。  $n = 10$  よりも棒の数が増えるものの、棒の縦幅 (確率) が小さくなっていて、結果として棒の面積の合計は変わらず  $\underline{\hspace{1cm}}$ 。



- $n = 300$  のとき、期待値  $E(X)$  は  $E(X) = \underline{\hspace{1cm}}$  だから、大体 50 回前後の確率が大きくなっている。棒はさらに細分化され、棒それぞれの縦幅 (確率) はとても小さくなり、棒全体の面積の合計は  $\underline{\hspace{1cm}}$ 。さらに、 $P(50 \leq X \leq 70) \simeq \frac{1}{2}$  も見て取れる。(← 全体「1」に対する割合と理解できる。)



- 確率密度関数  $f(x)$  とは、 $n$  を十分大きくしたときに現れる曲線  $y = f(x)$  を表す関数であると理解でき、曲線  $y = f(x)$  の表す山全体の面積は  $\underline{\hspace{1cm}}$  である。
- そして確率  $P(a \leq X \leq b)$  とは、全体に対して、範囲  $a \leq X \leq b$  がどのくらいの割合を占めているのか、つまり、確率がどの程度詰まっているのかを表しているものと理解できる。

まとめ

確率密度関数  $f(x)$  とは、文字通り確率の密度を表す関数であり、グラフと横軸が囲う箇所の面積<sup>a</sup>は 1 である。全体を「1」にすることで、表す範囲の面積が、割合、すなわち確率になる。

<sup>a</sup> 教科書にある記号  $\int_a^b$  は積分の記号で、面積を表すものである。使わなくても理解できるからここではふれない。

♠ エキスパート B 「正規分布」

目標 B

正規分布の定義の意味を大雑把に理解し、グラフと式の関係を説明できる。

説明

定義.  $m$  を実数、 $\sigma$  を正の実数とする、確率変数  $X$  の確率密度関数  $f(x)$  が

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

であるとき、確率変数  $X$  は正規分布  $N(m, \sigma^2)$  に従うといい、

$$X \sim N(m, \sigma^2)$$

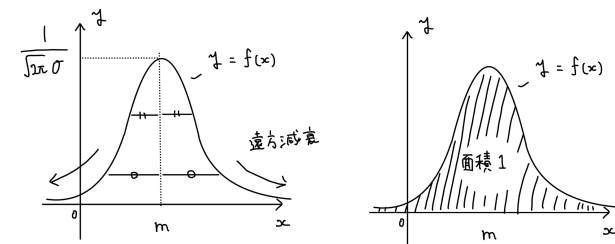
と書く。ただし、 $e$  とはおおよそ  $e \simeq 2.71828$  を満たす無理数である。

- $e$  は  $\pi$  や  $\sqrt{2}$  と同じ、ただの無理数。詳しい性質は数学 3 で習う。
- 指数部分の分子  $-(x-m)^2$  は  $x$  の値によらずに常に【 0 以上 / 0 以下 】であり、 $x = \underline{\hspace{1cm}}$  のときに最大値を  $\underline{\hspace{1cm}}$  をとる。
- $f(m+\sigma)$  と  $f(m-\sigma)$  の値が【 等しい / 等しくない 】ことからわかるように、 $y = f(x)$  のグラフは、 $x = m$  に関して【 対称である / 対称ではない 】。← 期待値 (山の中心) が  $m$  になるように調整されている。
- 指数部分の分母は  $2\sigma^2 \leftarrow$  標準偏差 (山の散らばり) が  $\sigma$  になるように調整されている。
- $x$  の値が  $x = m$  から離れていくと、

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \underbrace{\frac{1}{e^{\frac{(x-m)^2}{2\sigma^2}}}}_{\text{マイナス乗は逆数}} \xrightarrow{x \text{ を遠方へ}} 0$$

となり、 $f(x)$  の値は【 どんどん小さく / どんどん大きく 】0 に近づいていく。← グラフの遠方減衰性。

- 定数倍  $\frac{1}{\sqrt{2\pi}\sigma}$  は山全体と  $x$  軸が囲う場所の面積が 1 になるように調整された数である。



- そして不思議なことに、自然現象 (個体差 (身長、体重) など) や社会現象 (テストの得点など) の分布が、正規分布に近い分布になることが知られている (詳しくは大学で)。

まとめ

正規分布の定義は、期待値と標準偏差がそれぞれ  $E(X) = m$ 、 $\sigma(X) = \sigma$  となるように調整された定義であると理解せよ。