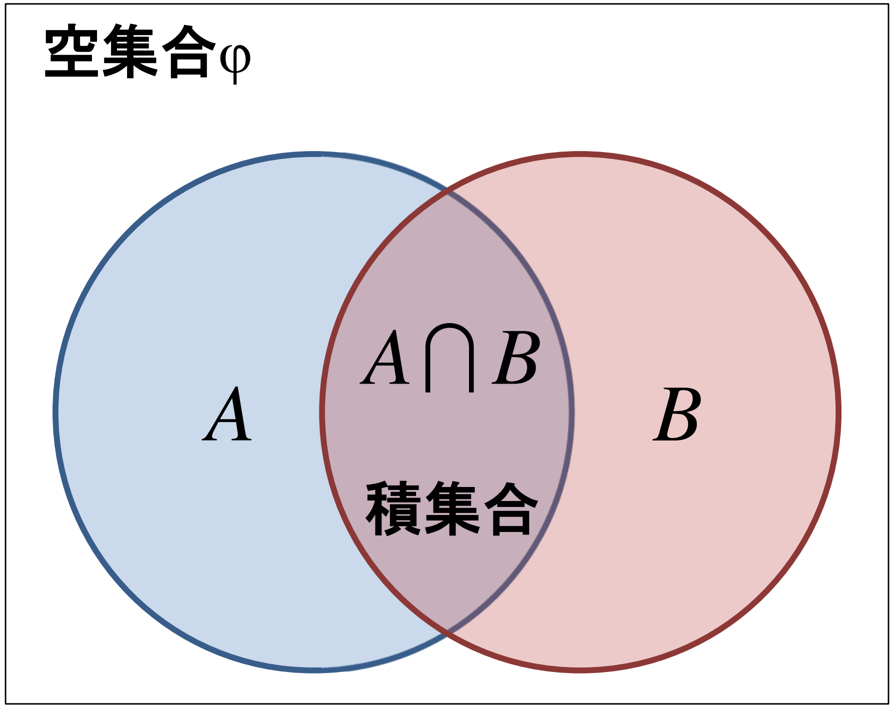
統計学入門
～確率変数と確率分布～

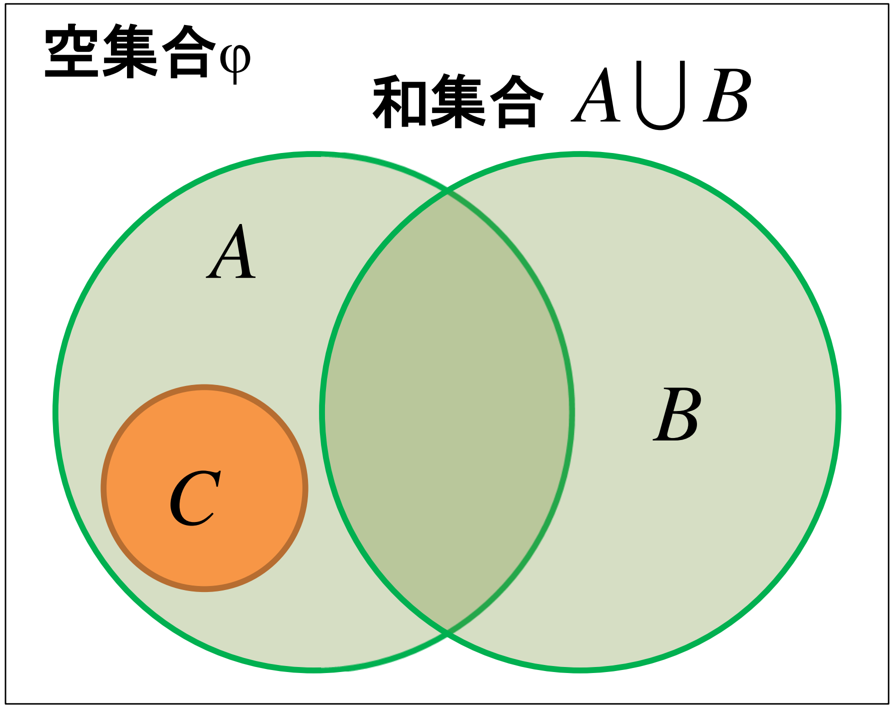
2025年度1学期：　月曜2限

担当教員：　石垣　司

集合

和集合, 積集合, 空集合, 部分集合





部分集合 $C \subseteq A$

※今回の講義内容のほとんどは高校数学の復習

用語の整理 #1

試行：結果が偶然に支配される実験

- 例：サイコロを転がす実験

根元事象：試行の結果の素となる事象

- 例：サイコロの試行で出る目の数
- $A_1 = \{1\}, A_2 = \{2\}, \dots, A_6 = \{6\}$

標本空間：すべての根元事象を集めた集合

- 例：サイコロは1から6の目が出る
- $S = \{A_1, A_2, A_3, A_4, A_5, A_6\} = \{1,2,3,4,5,6\}$

用語の整理 #2

事象：標本空間の任意の部分空間

- **例**1：**偶数の目が出る**　$A = \{A_2, A_4, A_6\}$
- **例**2：4**以下の目が出る**　$B = \{A_1, A_2, A_3, A_4\}$
- **事象の数の表現** # (**例**：$\#A = 3, \#B = 4$)

空事象$\{\varphi\}$：**起こり得ない事象**

- **例**：6**面サイコロの試行で**8**が出る**

全事象：一番大きな事象

- **例**：$S = \{A_1, A_2, A_3, A_4, A_5, A_6\} = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$

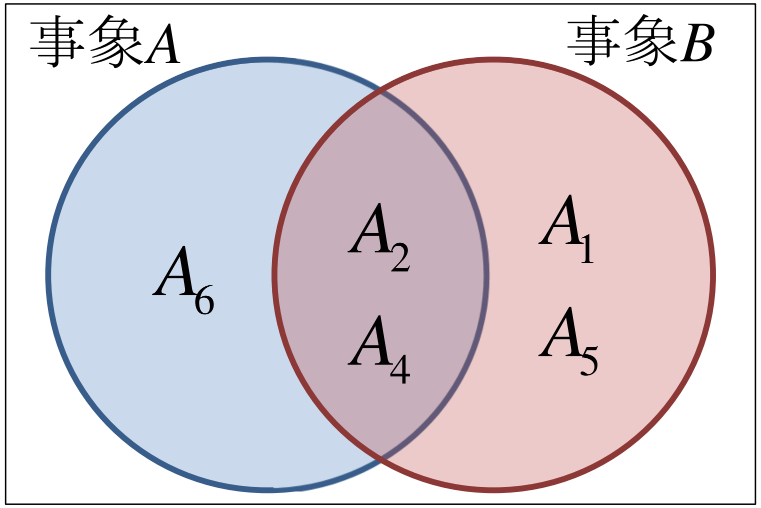
用語の整理 #3

和事象：2つの事象の和集合になる事象

- 例：「偶数」と「4以下」の和事象
- $A \cup B = \{A_2, A_4, A_6\} \cup \{A_1, A_2, A_3, A_4\} = \{A_1, A_2, A_3, A_4, A_6\}$

積事象：2つの事象に共通な事象

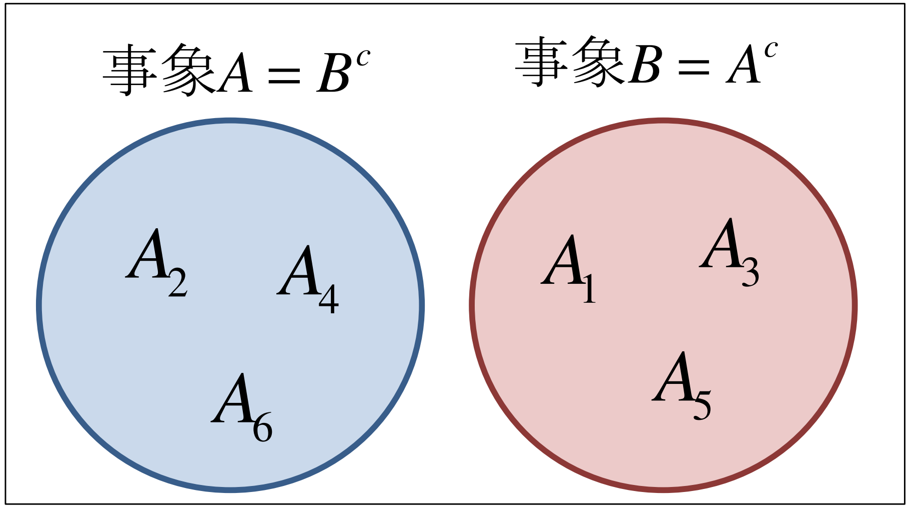
- 例：「偶数」と「4以下」の積事象
- $A \cap B = \{A_2, A_4, A_6\} \cap \{A_1, A_2, A_3, A_4\} = \{A_2, A_4\}$



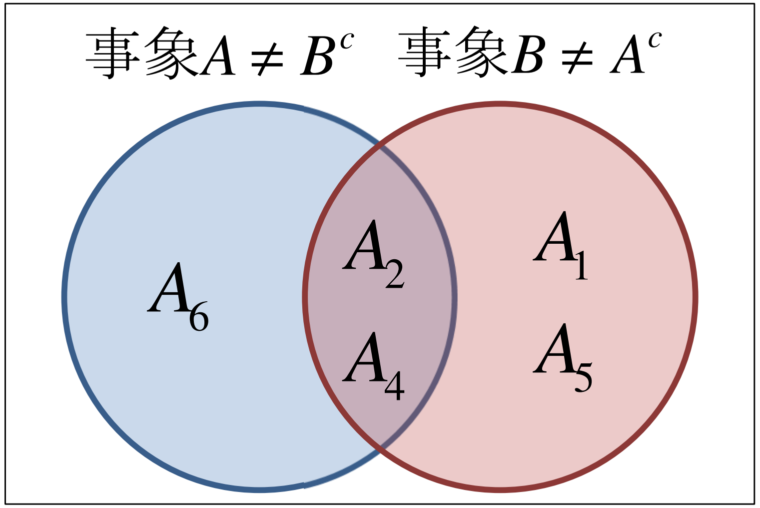
用語の整理 #4

排反事象：共通な事象をもたない2つの事象
- 排反事象の積事象は空事象
- 例1：$A = \{A_2, A_4, A_6\}$ と $C = \{A_1, A_3, A_5\}$ は排反
- 例2：$A = \{A_2, A_4, A_6\}$ と $B = \{A_1, A_2, A_3, A_4\}$ は排反ではない

余事象：事象 A が起きない事象 A^c
- 例：$A = \{A_2, A_4, A_6\}$ と $A^c = \{A_1, A_3, A_5\}$



例1



例2

用語の整理 #5

順列：n 個の種類の異なる対象から r 個を取り出して順に並べたときの内容

– 順列の総数

$$ {}_nP_r = \frac{n!}{(n-r)!} $$

組合せ：n 個の種類の異なる対象から選ばれた r 個の内容

– 組合せの総数

$$ {}_nC_r = \frac{{}_nP_r}{r!} $$

様々な確率の定義 #1

確率とは？

– 偶然性を伴う事象が生じる可能性の尺度の表現

古典的確率（数学的確率）

$$P(A) = \frac{\text{事象}A\text{に含まれる根源事象の数}}{\text{標本空間に含まれる根源事象の数}}$$

Pierre-Simon Laplace 1746-1827

統計的確率

– 試行の回数 n で 事象 A の生じる回数 r のとき

$$P(A) = \lim_{n\to\infty} \frac{r}{n}$$

主観的確率

– 信念の度合いを 0 ~ 1 の数値で表現

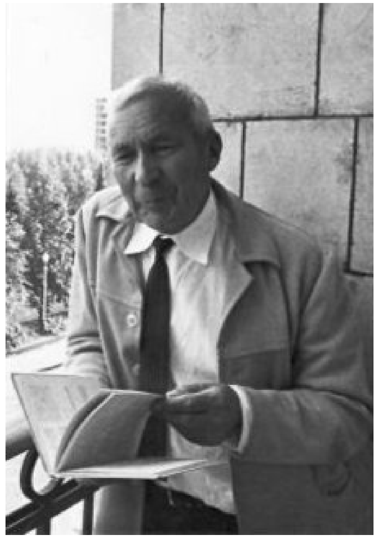
例：彼/彼女への告白が実る確率はせいぜい10 %程度だろう

例：その新規事業は8割以上の確率で成功するだろう

様々な確率の定義 #2

確率の公理 (公理　理論を展開する前提)

1. **任意の事象** A **に対し** $0 \leq P(A) \leq 1$
2. **全事象** S **に対し** $P(S) = 1$
3. $\underbrace{A \cap B = \varphi}_{A\text{ と }B\text{ が排反}}$ **ならば** $P(A \cup B) = P(A) + P(B)$



Andrey Kolmogorov 1903-1987

- **上の3つを満たす関数** P **を確率として扱う**
- **古典的・統計的・主観的な各確率は確率の公理を満たすように定義可能**

  ※コルモゴロフの公理(1933)の正確な理解には確率測度が必要。その内容は本講義では扱わない

確率の性質

余事象の確率: $P(A^c) = 1 - P(A)$

事象 A **が事象** B **の部分集合なら** $P(A) \leq P(B)$

加法定理: $P(A) + P(B) = P(A \cup B) + P(A \cap B)$

空事象の起こる確率 $P(\varphi) = 0$

同時確率と条件付き確率

同時確率：**事象 A と B が同時に生じる確率**

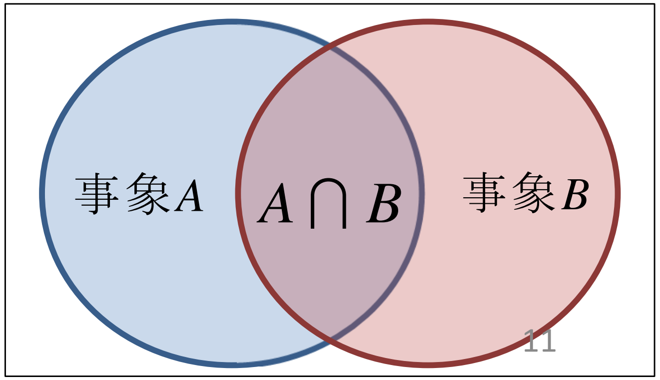
$$P(A,B) = P(A \cap B)$$

条件付き確率：**事象 A が生じた後で B が生じる確率**

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

確率の乗法定理

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



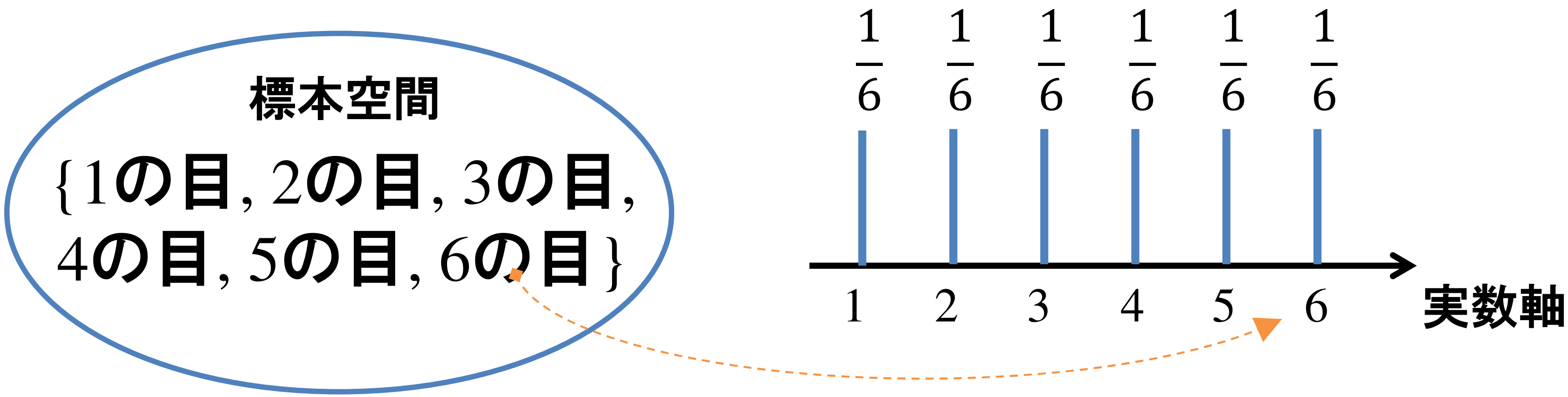
確率変数

- どのような値となるかが，ある確率法則によって決まる変数（日本工業規格）
- 事象を実数 x に結び付ける関数
  - 本講義では確率変数は大文字 X や Y で記述する
  - データ，実現値，実数値は x や y で記述する
    ※確率変数の数学的定義は測度論が必要。本講義では扱わない
- 離散型確率変数
  - 例：サイコロ，コイン投げ
- 連続型確率変数
  - 例：気温，身長，体重

離散型確率変数

標本空間の根元事象の数が整数の場合の確率変数

- 例：サイコロの確率変数　X (1の目) = 1, …, X (6の目) = 6
- 事象の生じる確率　$P(X=1)=\frac{1}{6},\ldots,P(X=6)=\frac{1}{6}$



X の確率分布

- 実数軸上の青い縦線は確率関数 $p_X(x_i)$

  縦線の上の数値は各事象の生じる確率

確率関数

確率関数(確率質量関数) p_X

- 確率変数がとる全ての値と, その値が生じる確率を記述した関数
- 離散型確率分布の表現
- 例：サイコロ　$p_X(1) = \frac{1}{6}, \dots, p_X(6) = \frac{1}{6}$　$(p_X(-1) = 0, p_X(8) = 0)$

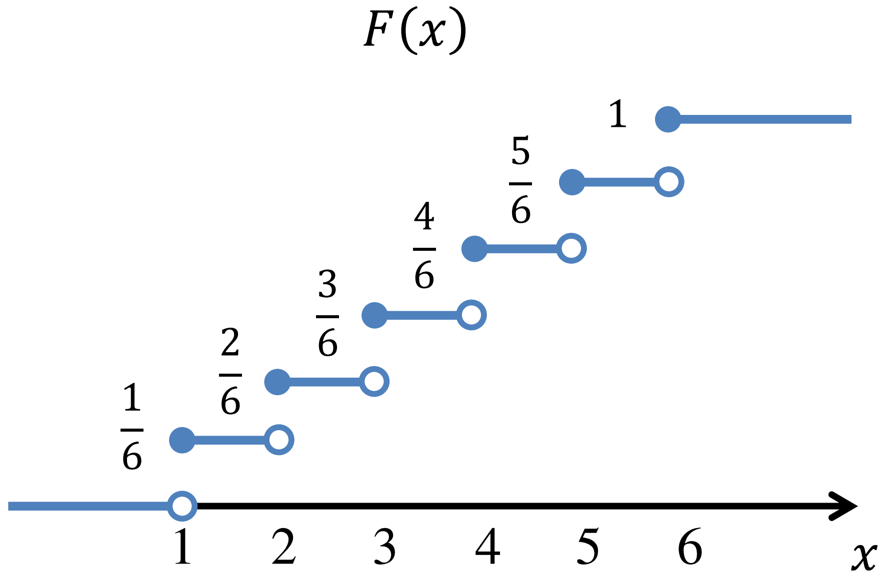
確率関数 p_X の定義

- 確率変数 X は m 個の実数値で次を満たす関数
- $p_X(x_1) = p_1, \dots, p_X(x_m) = p_m$　(p_i は正の確率)
- $x_1, \dots, x_m$ 以外の x では $p_X(x) = 0$
- $\sum_{i=1}^{m} p_i = p_1 +, \dots, + p_m = 1$

（離散）累積確率分布関数

分布関数(累積確率分布関数) $F(x)$

- $F(x) = P(X \leq x) = \sum_{i=1}^{j} p_X(x_i)$
- x_j は $x_1 < x_2, \dots, < x_m$ のうち, x 以下の最小値
- 例：サイコロ
- $F(x) = \begin{cases} 0, & x < 1 \\ 1/6, & 1 \leq x < 2 \\ 2/6, & 2 \leq x < 3 \\ 3/6, & 3 \leq x < 4 \\ 4/6, & 4 \leq x < 5 \\ 5/6, & 5 \leq x < 6 \\ 1, & 6 \leq x \end{cases}$



離散型確率分布の例

ベルヌーイ分布

Jakob Bernoulli
(1654-1705)

- 結果が2種類しかない試行に関する分布
  - 例 / コイン投げの{表, 裏}
  - 例 / 消費者の商品購買{購買,非購買}
- 標本空間　$S = \{0, 1\}$
- 確率　$P(X = 1) = p, P(X = 0) = 1 - p$
- パラメータ　p

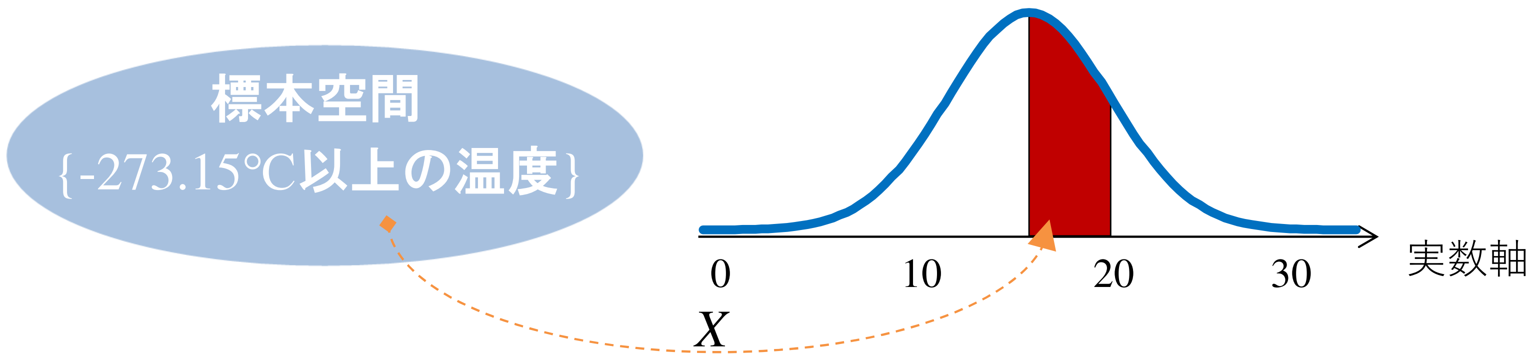
問題

- 試行の結果が x のときのベルヌーイ分布の確率関数は $P(X = x) = p^x(1 - p)^{1-x}$ と表現できる。このとき, ベルヌーイ分布の累積確率分布関数を求めなさい

連続型確率変数

標本空間が連続の確率変数

- **例：明日の最高気温　(確率変数** X)
- **事象がある範囲に入る確率**　$P(15℃ \leq X \leq 20℃)$



- **実数軸上の青い曲線が確率密度関数** $f(x)$
- **赤色部分の面積が確率** $P(15℃ \leq X \leq 20℃)$

確率密度関数

確率密度関数 $f(x)$

- 確率変数がとる全ての値の相対的な出現のしやすさを記述した関数
- 連続型確率分布の表現

連続型確率変数 X が範囲 $[a, b]$ に値をとる確率

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

※離散型確率変数と異なり $P(X = 15℃)$ は意味を持たない

$\because\ P(X = a) = 0$ （a は任意の実数）

確率密度関数の性質

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

(連続)累積確率分布関数

分布関数(累積確率分布関数) $F(x)$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

確率変数 X が x よりも小さい値をとる確率

– **分布関数と確率密度関数の関係**

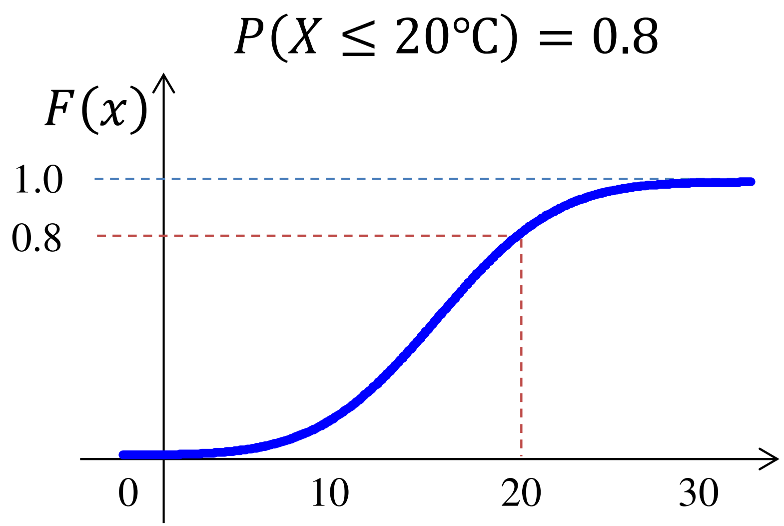
$$F'(x) = f(x)$$

累積確率分布関数の性質

– $F(x)$ **は単調増加関数**

– $0 \leq F(x) \leq 1$

– $F(-\infty) = 0, F(\infty) = 1$



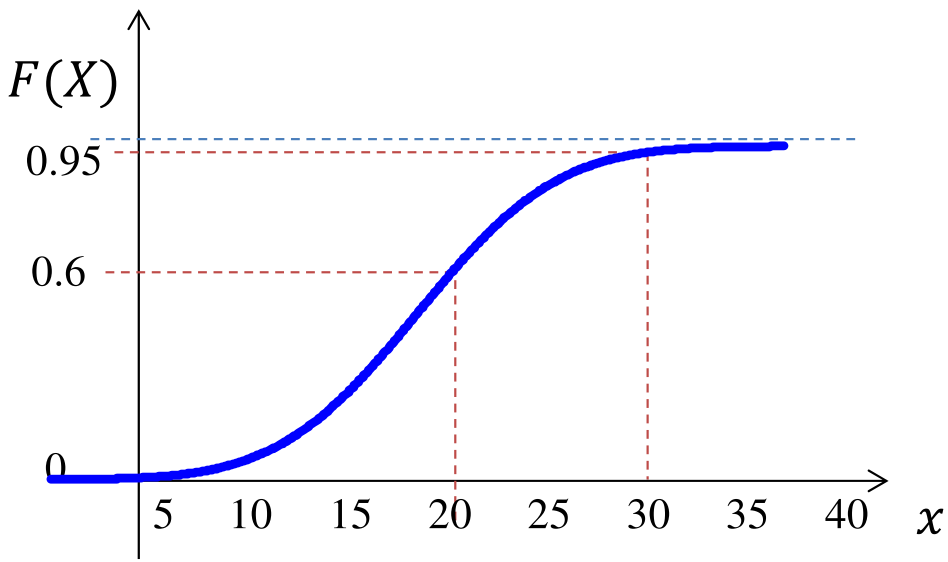
連続型確率変数の確率

確率変数 X が区間 $[a, b]$ に入る確率

- **累積確率分布関数による表現**

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

- **例題：明日の最高気温 X を考え，その累積分布関数が下図で表現されるとする。このとき，明日の最高気温が20℃以上，30℃以下となる確率を求めよ**



求め方：

$P(20.0 \leq X \leq 30.0)$

$= F(30.0) - F(20.0)$

$= 0.95 - 0.6$

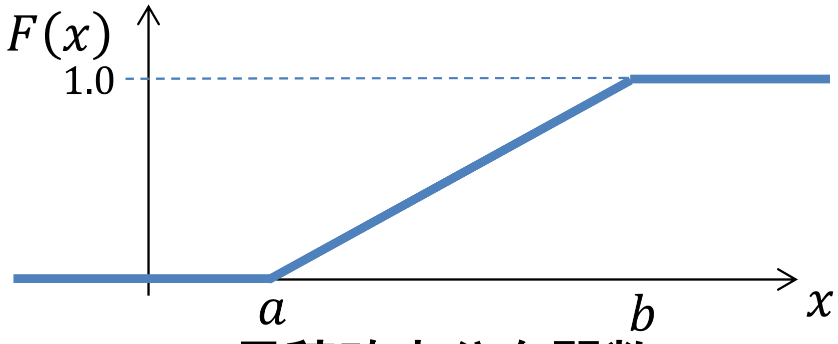
$= 0.35$

連続型確率分布の例

一様分布

- 区間 $[a, b]$ で密度関数が一定の分布
- 確率密度関数　$a \le x \le b$ のとき $f(x) = \frac{1}{b-a}$

確率密度関数



累積確率分布関数

問題

1. 一様分布の累積確率分布関数を求めなさい
2. 確率変数 X は区間 $[1,3]$ の一様分布に従う。このとき，$P(1 \le X \le 2)$ を求めなさい

連続型確率変数の同時・条件付き確率

例：期間 $[0,\infty]$ のある時刻 X に機械が故障する確率

- **期間 $[a,b]$ の間に機械が故障する確率**

$$P(a \leq X \leq b)$$

- **時刻 $c(c < a)$ まで故障が起こらなかったとき（つまり，時刻 c 以降に故障が生じる $(c \leq X)$）に，時刻 c 以降の期間 $[a,b]$ の間に機械が故障する確率**

$$P(a \leq X \leq b | c \leq X)$$

問題

- **条件付き確率 $P(a \leq X \leq b | c \leq X)$ を $P(a \leq X \leq b)$ と $P(c \leq X)$ で表現しなさい**

演習問題

問題

– 確率密度関数を

$$f(x) = \begin{cases} -6x^2 + 6ax & (0 \leq x \leq a) \\ 0 & (x < 0, a < x) \end{cases}$$

とするとき，a の値と累積分布関数 $F(x)$ を求めよ

ヒント $\int_{-\infty}^{\infty} f(x)dx = \int_0^a f(x)dx = 1$