序論：確率論と統計学
確率空間

- 序論：確率論と統計学
  - 確率的現象
  - 確率論
  - 統計学
  - 本講義の予定
- 確率空間
  - 確率現象のモデル化
  - 素朴な確率空間
  - 公理的な確率空間
  - 条件付き確率と事象の独立性

確率論と統計学

（1）確率的な現象

- この世のほとんどの現象は「確率的」:
  - くじ，ギャンブル，各種スポーツ
  - 金融市場での証券（株式，債券等）価格，金利，為替
  - 人の出会い，交通事故，機械の故障，病気の発症
  - 天候，自然災害（洪水，地震） *etc. etc…* （列挙しだすときりがない）

- 土木系3コース分野での「確率的現象」の例
  - 社会基盤デザイン系:

    橋梁への荷重，コンクリート材料/地盤の強度，地震等の外力 *etc.*
  - 水環境デザイン系:

    河川流域の降水量/流量，汚染物質の拡散，感染症の流行 *etc.*
  - 都市システム計画系:

    交通施設の需要（交通量），消費者の選択行動，地価の変動 *etc.*

確率論と統計学

（2）確率論 – その1

- 「確率論（*Probability Theory*）」とは：

  「確率的 (*Probabilistic/Stochastic/Random*) 」現象を数学的に「**モデル化**」し，その確率法則のもとで，**確率的変数・システムが持つ性質や規則性** を（定量的／定性的に）解明する

- 「**確率論的**」現象と「**確定論的**」現象の不思議 (?) な関係
  - コインを1回だけ投げた時に表／裏のどちらが出るか？
    ・・・ まったく**予測不可能**（「確率的」 現象）
  - コインを1000回投げた時に表が出る相対頻度は？
    ・・・「平均的」には，ある程度，予測可能（でも，かなりバラつく）
  - コインを10万回投げた時に表が出る相対頻度は？
    ・・・ **非常に高い精度で予測可能 !!!**（「確定論的」 現象）

確率論と統計学

（2）確率論 – その2

- 「確率論」の身近な応用例：
  - *Amazon.com* の “おすすめサービス” （「**ベイズの定理**」）
  - *Google* の“*PageRank*” （「**マルコフ連鎖**」）
  - 各種の保険 （「**大数の法則／中心極限定理**」）
  - 年金運用（証券投資）のリスク管理 （「**確率変数の相関**」）

  . . . 等々.

- 「確率論」を基盤とする様々な理論

  統計学， 意思決定理論， ゲーム理論， 情報理論，
  信頼性理論， 品質管理論， リスク管理論，
  ファイナンス理論（投資理論，金融工学）， 在庫理論，
  待ち行列理論， 選択行動理論， 水文学， 耐震工学，
  画像処理理論， 機械学習理論， 制御理論， 統計力学

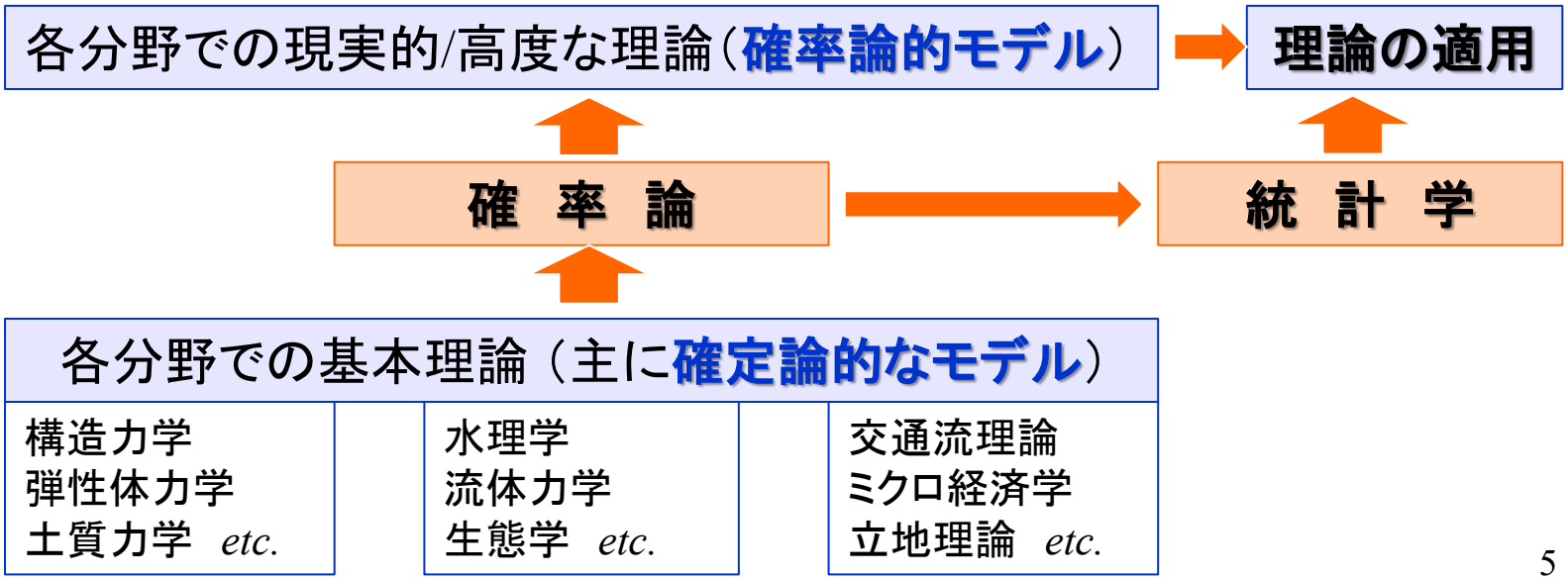
  . . . 等々

確率論と統計学

（3）確率論と統計学

- 「統計学（*Statistics*）／統計的分析」とは：
  ある確率法則（“確率モデル”）に従うと考えられる確率変数の**実現値**（“**データ**”）**を観測し**，それを生成する確率法則に関する
  なんらかの**推測**（**モデルの構造やパラメータの推定**）をする．

- 「確率論・統計学」と各専門科目の関係



確率論と統計学

（4）本講義の予定

赤松

（http://www.plan.civil.tohoku.ac.jp/~akamatsu/AppPrbStt/syllabus.html）

I. 確率論

第１回～第６回 （確率空間， 確率変数と確率分布， 期待値と特性関数，同時確率と共分散，条件付き確率とベイズの定理，大数の法則と極限定理）

II. 確率過程入門

第７回 （マルコフ連鎖の基礎と応用，待ち行列理論）

原

III. 統計的推論

第８回～第11回 （記述/推測統計学・データ可視化，点推定・信頼区間仮説検定，統計的推論と統計モデリング）

IV. 統計モデルと推定法

第12回～第14回 （線形/ロジスティック回帰モデル・最小二乗法と最尤推定モデル選択とAIC，ベイズ推定とMCMC）

第15回 (1/27) : 期末試験（赤松＋原）

確率空間

（1）確率現象のモデル化と確率空間

- 確率現象のモデル化
  - 確率論における「確率的現象」： 起こりうる複数の事象のうち，何れか1つの事象が（何らかの「偶然」によって）現実に生起する.
  - モデル化のために，

(1) 観測で区別できる事象の最小単位（**根元事象** *or* **標本点** ω）および，それら全てを集めた集合（**標本空間** Ω）を設定.

例： 1個のサイコロ振り：$\Omega = \{1,2,3,4,5,6\}$ （Ω の各要素が ω）

(2) 生起したかどうかを区別できる（単一または複数の標本点からなる）ことがら（**事象** E）を設定.

例： 1個のサイコロ振りで偶数が出る事象 $E = \{2,4,6\}$

(3) 各事象に対して，その「生起しやすさ」を表す数値（**確率**）を与える方式（関数 $P(E)$）を設定.

・・・・・ (1)～(3) をまとめて「**確率空間**」の設定と呼ぶ.

確率空間

（1）確率現象のモデル化と確率空間（おまけの復習）

- 「集合」に関わる初歩的な概念と表記
  - 事象 E は標本空間 Ω の**部分集合**： $E \subset \Omega$
  - **空集合**（起こらない事象）： $\varnothing$（*null*）
  - 事象 E の**余事象**（Ω の中で E が起こらない事象）： $E^C \equiv \Omega - E$
  - 事象 A と B の**和事象**（事象 A or B が起こる事象）： $A \cup B$
  - 事象 A と B の**積事象**（事象 A and B が同時に起こる事象）： $A \cap B$
  - 事象 A と B が**排反**（事象 A and B が同時に起こらない）： $A \cap B = \varnothing$
  - 集合 Ω が**有限集合**： Ω の要素数が有限.

    可算集合： Ω の要素数が有限でなく, Ω から自然数全体の集合へ1対1 で上への写像が作れる.

    非可算集合： 有限でも可算でもない集合（例：実数全体の集合）.
  - 集合 Ω の**ベキ集合** 2^{Ω} ：Ω の部分集合全体の作る集合族

確率空間

（2）素朴な確率空間 – その1

- 組合せ論的確率
  - **有限個**の根元的事象があり，各々が**等確率で生起**する場合．

    事象 E の生起確率：$P(E) = |E| / |\Omega|$ （$|E|$ は E の**要素数**）

    例1： 1個のサイコロ振りで偶数の目が出る確率

    $\Omega = \{1,2,3,4,5,6\},\ |\Omega| = 6,\ E = \{2,4,6\},\ |E| = 3$

    $\Rightarrow P(E) = |E| / |\Omega| = 3/6 = 1/2$

    例2： 丁半賭博で「丁」（2個の目の合計が偶数）が出る確率

    $\Omega = \{$省略$\},\ |\Omega| = 6 \times 6 = 36,\ E = \{$省略$\},\ |E| = 3 \times 6 = 18$

    $\Rightarrow P(E) = |E| / |\Omega| = 18/36 = 1/2$

  - 確率 $P(E)$ の性質：

    (i) $P(A) \geq 0 \quad \forall A \in S$ (ii) $P(\Omega) = 1,\ P(null) = 0$

    (iii) A, B が互いに排反事象なら $P(A \cup B) = P(A) + P(B)$

確率空間

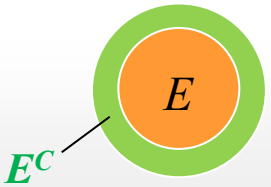
（2）素朴な確率空間 – その2

- 幾何学的確率
  - Ω が**無限個**の根元的事象からなる場合（**例**：連続的な実数空間）
    ⇒ 組合せ論的確率と同様に各根元的事象 ω が等確率 p で生起すると考えると，任意の要素数 n をとりうる事象 E の確率が $0 \le P(E) = pn \le 1$ を満たす様に根元的事象の確率 p を決めることができない．．．
  - 事象 E の生起確率：$P(E) = |E| / |\Omega|$ （$|E|$ は E の**面積**）

例：射撃の標的 Ω が半径 3 の円のとき，中心が Ω と同一で半径 2 の円内領域 E に命中しない（*i.e.*, 幅 1 の周縁部 E^C に命中する）確率：

$|\Omega| = 9\pi,\ |E| = 4\pi,\ |E^C| = |\Omega| - |E| = 5\pi$

$\Rightarrow P(E^C) = |E^C| / |\Omega| = 5/9$



- 確率 $P(E)$ の性質：

(i) $P(A) \ge 0 \quad \forall A \in S$ (ii) $P(\Omega) = 1,\ P(null) = 0$

(iii) A, B が互いに排反事象なら $P(A \cup B) = P(A) + P(B)$

確率空間

（3）公理論的な確率空間 – その1

- 事象の集合（可算加法族 または σ-algebra）
  - **定義**：S が集合 Ω の部分集合の族で，以下の性質を満たすとき，S を Ω 上の**「可算加法族」**（**事象**の集合）と呼ぶ：

(i) $\Omega \in S$ （Ω 自身が事象）

(ii) $A \in S \Rightarrow A^C \in S$ （A が事象なら A の**補集合** A^c も事象）

(iii) $A_1, A_2, \ldots \in S \Rightarrow \bigcup_{k=1}^{\infty} A_k \in S$ （A_k が事象なら A_k の**和集合**も事象）

- 確率
  - **定義**（**Kolmogorovの公理**）：集合 Ω 上の可算加法族 S を定義域とし，以下の性質を持つ関数 P を S 上の**確率**（**確率測度**）と呼ぶ：

(i) $P(A) \geq 0 \quad \forall A \in S$ (ii) $P(\Omega) = 1$

(iii) $A_1, A_2, \ldots \in S,\ A_j \cap A_k = null \ (j \neq k)$ （$A_1, A_2, \ldots$ が互いに**排反事象**）

$$\Rightarrow P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

確率空間

（3）公理論的な確率空間 – その2

- 確率空間
  - **定義**: S をΩ上の可算加法族（事象の集合），P を S 上の確率とする．これらの3つ組(Ω, S, P)を**確率空間** (*probability space*) と呼ぶ．Ω を**標本空間** (*sample space*)，Ω の要素を**根元事象** (*elementary event*) または標本点 (*sample point*)，S の要素を**事象** (*event*) と呼ぶ．

- 確率の基本的性質
  - **定理**: (Ω, S, P)を確率空間とし，$A, B \in S$ とする．このとき，

$$(1)\quad P(A^c) = 1 - P(A)$$

$$(2)\quad B \subset A \Rightarrow P(A - B) = P(A) - P(B)$$

$$(3)\quad B \subset A \Rightarrow P(B) \leq P(A)$$

$$(4)\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

確率空間

（3）公理論的な確率空間 – その3

- **例**： 2回のコイン投げ
  - 根元事象: $\omega_1 = \{表表\},\ \omega_2 = \{表裏\},\ \omega_3 = \{裏表\},\ \omega_4 = \{裏裏\}$
  - 標本空間: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
  - 各根元事象の確率: $P(\omega_k) = 1/4\ \ (k = 1,2,3,4)$
  - 事象 $A = \{1$回目に表$\}$, $B = \{2$回目に表$\}$ とする.

$$\Rightarrow \quad A = \{\omega_1, \omega_2\},\ B = \{\omega_1, \omega_3\},\ A \cap B = \{\omega_1\}$$

$$P(A) = 1/2,\ P(B) = 1/2,\ P(A \cap B) = 1/4$$

$$\Rightarrow \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/4$$

- **問**： 無限反復のコイン投げ
  - 表が出るまでコインを投げ続け，根元事象$\omega_k = \{k$ 回目に初めて表$\}$ $(k = 1,2,\ldots)$とする（標本空間は**可算集合**）.
  - n 回以上投げ続ける事象をA_n とする. コルモゴロフの公理を満たす $P(A_n)$ を示せ.

確率空間

（3）公理論的な確率空間 – その4（参考補足）

- 実数全体の集合 R 上の確率空間
  - $\Omega = R = (-\infty, +\infty)$ の場合，確率が定義・計算できる事象集合 S のとりかたは自明ではない（「**ルベーグ積分**」の高度な議論が必要）．
  - 事象集合 S の一つの標準的な設定は，「**ボレル集合族**」$B(R)$（R の全ての半開区間 $(a, b]$ を含む最小の可算加法族）．
  - 事象集合 $B(R)$ に対して Kolmogorov公理で定義される確率 P は，より具体的には，以下の条件を満たす**分布関数** $F(x) = P((-\infty, x])$ によって構成できる：

$$\text{(i)} \lim_{x \to -\infty} F(x) = 0, \qquad \text{(iii)}\ x < y \Rightarrow F(x) < F(y)\ \text{（単調増加）}$$

$$\text{(ii)} \lim_{x \to +\infty} F(x) = 1, \qquad \text{(iv)} \lim_{\varepsilon \downarrow 0} F(x+\varepsilon) = F(x)\quad \text{（右半連続）}$$

  - 分布関数 $F(x)$ に対して，$F(x) = \int_{-\infty}^{x} f(u)du$ を満たす関数 f を**密度関数**と呼ぶ．

確率空間

（4）条件付き確率と事象の独立性 – その1

- **条件付き確率**
  - **定義**：$P(B)>0$ のとき，

$$P(A\mid B)=P(A\cap B)/P(B)$$

を事象 B が起こった場合の事象 A の**条件付き確率** と呼ぶ．

  - **例**：2回のサイコロ投げで，$\Omega=\{(i,j)\mid i,j=1,2,...,6\}$，

$A=\{$少なくとも1回，2 が出る$\}$，$B=\{$目の和が6$\}$ とする．

$\Rightarrow B=\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$，　$A\cap B=\{(2,4),(4,2)\}$

$\Rightarrow P(B)=\frac{5}{36}$，　$P(A\cap B)=\frac{2}{36}$ $\Rightarrow \underline{P(A\mid B)=(\frac{2}{36})/(\frac{5}{36})=2/5}$

- **条件付き確率の基本公式**
  - 定義より，

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

  - $\Omega=\bigcup_{k=1}^{n}B_k$ かつ B_k が排反である時，任意の事象 A に対して

$$P(A)=\sum_{k=1}^{n}P(A\cap B_k)=\sum_{k=1}^{n}P(A\mid B_k)P(B_k)$$

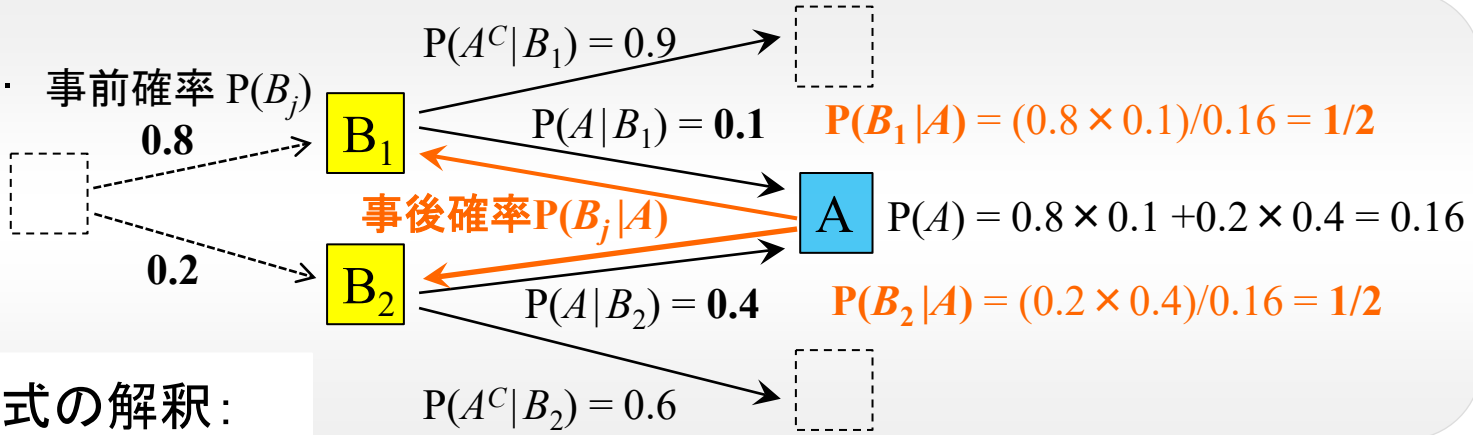
確率空間

（4）条件付き確率と事象の独立性 – その2

- **ベイズ（Bayes）の公式**
  - $\Omega = \bigcup_{k=1}^{n} B_k$ かつ B_k が排反である時，任意の事象 A に対して

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{k=1}^{n} P(A \mid B_k)P(B_k)}$$

  - **例**.



  - 公式の解釈：

事象 A が，事象 $\{B_i\}$ の何れか1つを「原因」とする，「結果」とする．原因事象 B_j が生起する（**先験的**／**事前の**）確率 $P(B_j)$ は，結果 A が起こったという情報によって，**事後的な**（その結果生起を条件とする）B_j の生起確率 $P(B_j \mid A)$ に修正される．

確率空間

（4）条件付き確率と事象の独立性 – その3

- **事象の独立性**
  - **定義**：事象 A, B が**独立** $\Leftrightarrow$ $P(A \cap B) = P(A)P(B)$

    $\Leftrightarrow P(A \mid B) = P(A), \quad P(B \mid A) = P(B)$

    （*i.e.*, 事象A, B の生起確率が互いの生起に影響されない）
  - **定義**：事象 $A_k \ (k = 1,2,\ldots,n)$ が独立である $\Leftrightarrow$ そこから選んだ任意の $m \ (1 \le m \le n)$個の事象 $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ に対して

    $$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_m})$$

例：2回のコイン投げで，A_1={表表，表裏}，A_2={表裏，裏裏}，A_3={表表，裏裏} とする．このとき，$P(A_k) = 1/2 \ (k = 1,2,3)$

$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = 1/4 \Rightarrow$ 各ペアは独立

$P(A_1 \cap A_2 \cap A_3) = 0 \Rightarrow (A_1, A_2, A_3)$ は独立ではない．

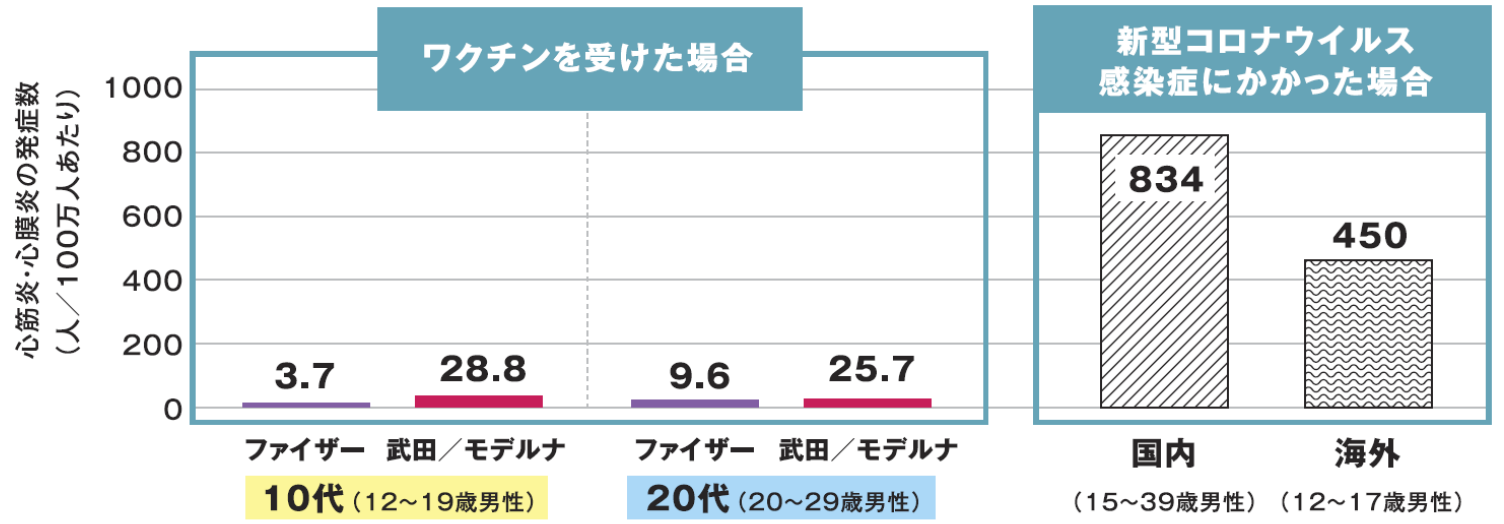
確率空間

おまけ：条件付き確率の（初歩的）応用

■10代・20代の男性も、引き続きワクチンの接種をご検討ください。

■10代・20代の男性は、ファイザー社のワクチンの接種も選択できます。

〈 心筋炎・心膜炎が疑われた報告頻度の比較（男性）〉



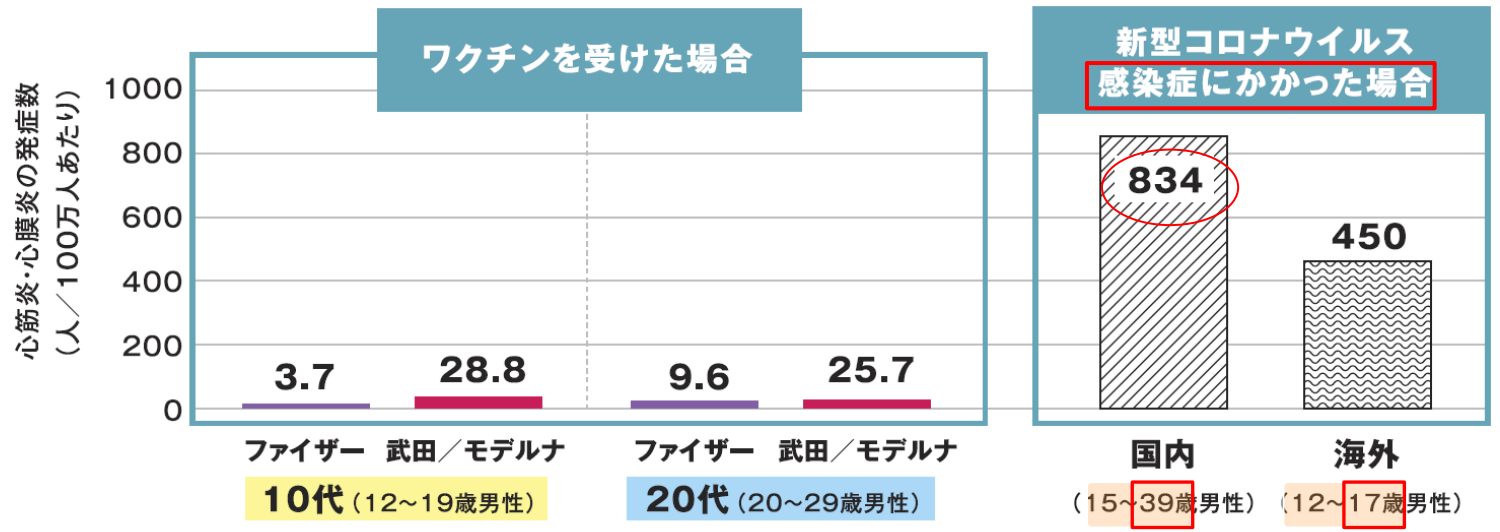
出典：第70回厚生科学審議会予防接種・ワクチン分科会副反応検討部会、令和3年度第19回薬事・食品衛生審議会薬事分科会医薬品等安全対策部会安全対策調査会（令和3年10月15日開催）資料

厚生科学審議会 (予防接種・ワクチン分科会 副反応検討部会)
2021年10月15日資料1-1-1 (pp.11, 22, 30参照)
https://www.mhlw.go.jp/content/10601000/000844075.pdf

確率空間

おまけ：条件付き確率の（初歩的）応用

〈 心筋炎・心膜炎が疑われた報告頻度の比較（男性）〉



出典：第70回厚生科学審議会予防接種・ワクチン分科会副反応検討部会、令和3年度第19回薬事・食品衛生審議会薬事分科会医薬品等安全対策部会安全対策調査会（令和3年10月15日開催）資料

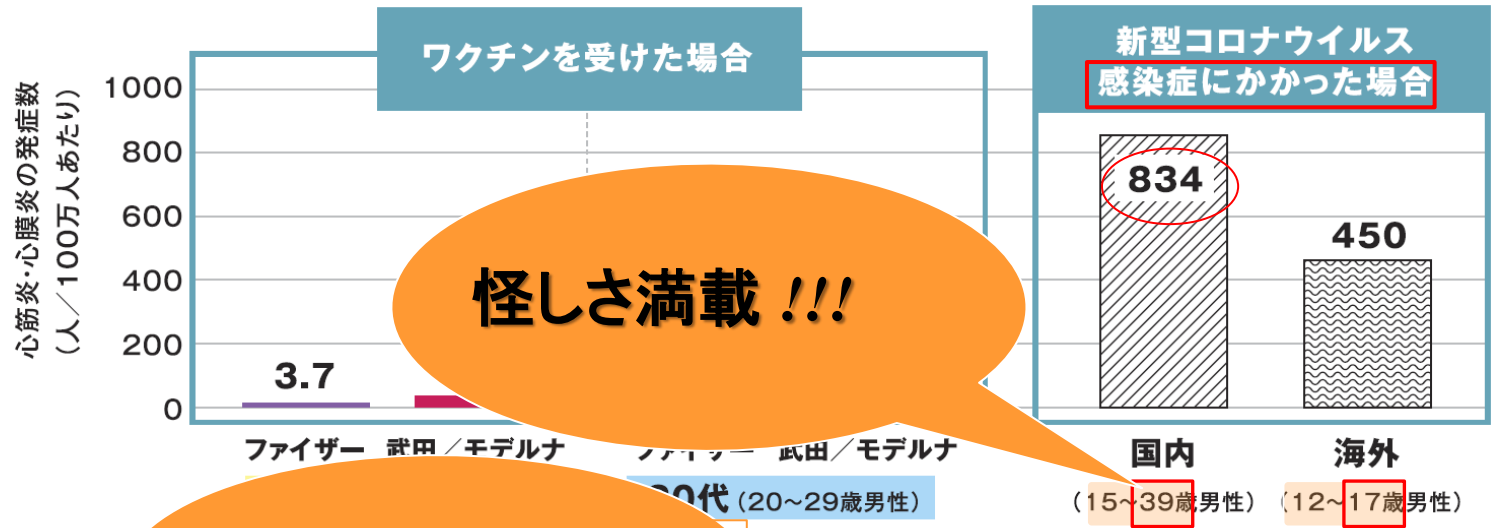
○レジストリー概要：新型コロナウイルス感染症と診断され、入院した症例。2021年5月31日時

	合計		15歳～40歳未満		40歳～65歳未満		65歳以上	
	男	女	男	女	男	女	男	女
対象人数	21,950	16,482	4,798	3,765	8,825	4,675	8,327	8,042
心筋炎関連事象者数*	23(4)	10(1)	4(1)	0	8(1)	5(0)	11(2)	5(1)
感染者100万人あたりの心筋炎関連事象者数	1,048	607	834	0	907	1,070	1,321	622

確率空間

おまけ：条件付き確率の（初歩的）応用

〈 心筋炎・心膜炎が疑われた報告頻度の比較（男性）〉



怪しさ満載 !!!

「詐欺商法」では !?

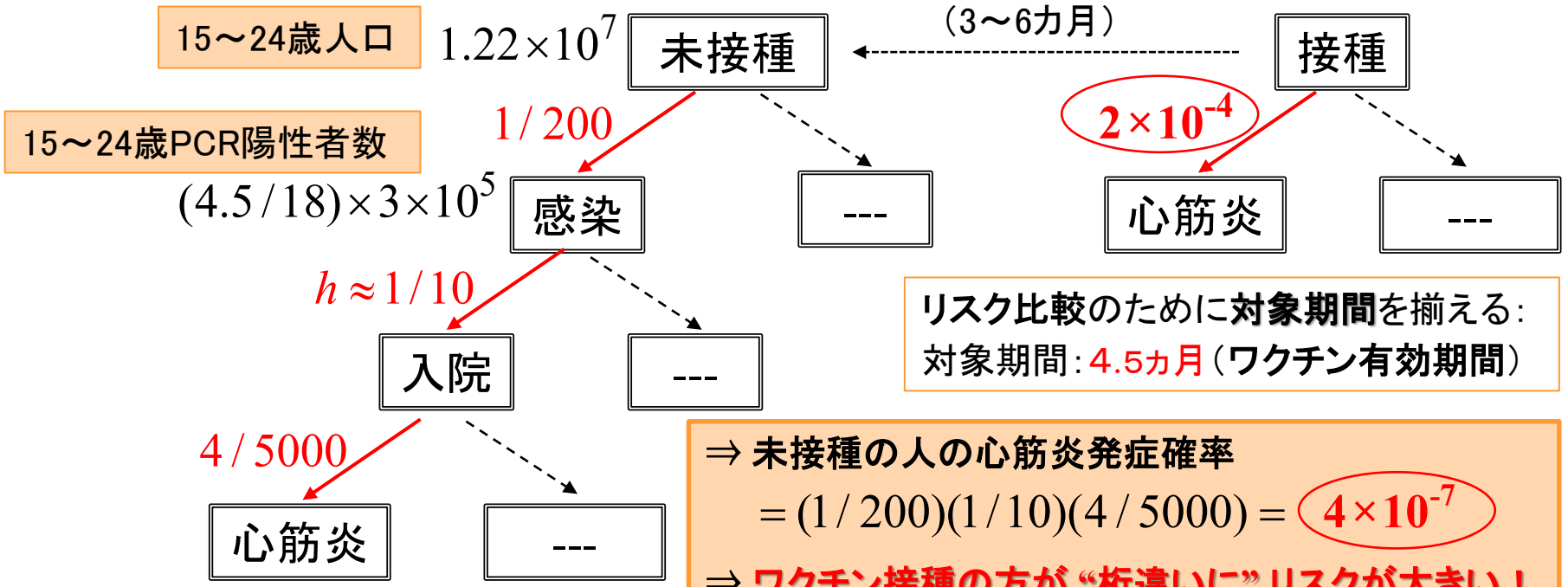
○レジストリ　　染症と診断され、入院した症例。2021年5月31日時

	合計		15歳～40歳未満		40歳～65歳未満		65歳以上	
	男	女	男	女	男	女	男	女
対象人数	21,950	16,482	4,798	3,765	8,825	4,675	8,327	8,042
心筋炎関連事象者数*	23(4)	10(1)	4(1)	0	8(1)	5(0)	11(2)	5(1)
感染者100万人あたりの心筋炎関連事象者数	1,048	607	834	0	907	1,070	1,321	622

確率空間

おまけ：条件付き確率の（初歩的）応用

- **心筋炎・心膜炎発症率**
  - Pfizer ワクチンを接種した人： 1/3000～1/6000 （15～24歳男性）
  - Covid-19に自然感染し入院した人： 4/5000 （15～39歳男性）
- **ワクチン接種と未接種のどちらのリスクが高いか？**



リスク比較のために**対象期間**を揃える：
対象期間：4.5カ月（**ワクチン有効期間**）

⇒ **未接種の人の心筋炎発症確率**

$$=(1/200)(1/10)(4/5000)=4\times10^{-7}$$

⇒ **ワクチン接種の方が “桁違いに” リスクが大きい！**

確率空間

おまけのおまけ：前頁の数値の根拠資料

- **Pfizer ワクチンを接種した若年男性の「心筋炎発症率」の根拠論文**
  - https://doi.org/10.1126/science.abj7796
  - https://doi.org/10.1056/NEJMoa2109730　https://doi.org/10.1056/NEJMoa2110737

  （以下の解説資料も読んでみてください：https://bit.ly/3k3gbgj）

- **15～24歳PCR陽性者の入院率 h の推計**

年齢区分	PCR陽性 (A)	男性入院	女性入院	入院計 (B)
15-39	785,018	4,798	3,765	8,563
40-64	522,945	8,825	4,675	13,500
65-	202,304	8,327	8,042	16,369

厚労省の入院者データが詳細年齢区分となっていないため，15～24歳と15～39歳の入院率が同じと仮定（本来は，若いほど低いが…）

- 仮定：**年齢別入院者数 ＝ 年齢別PCR陽性者数 × 年齢別入院率 × 標本率（年齢によらない）**
- 仮に，PCR陽性高齢者の入院率が（実際には1未満だが）**上限値** 1 とすると，

年齢区分	標本率	入院率	B / A
15-39	0.081	0.135	0.011
40-64	0.081	0.319	0.026
65-	0.081	1	0.081

15～39歳の入院率**上限値**は 0.135

$\Rightarrow 0.135 > h \approx 1/10$

（実際には，この値の数分の一未満でしょう）

確率空間

おまけのおまけ：念のための解説

- **厚労省の接種推奨チラシ（講義スライド p.20 のグラフ）には，以下の (1)〜(4) の「優良誤認表示」が施されており，本来のリスク評価結果（ワクチンのリスク >>> Covidのリスク）を真逆に捻じ曲げています：**

(1) ワクチン**接種する／しないの選択** に必要な（心筋炎発症リスクに関する）情報を提供するなら，本来，以下の2つの確率：

a) **接種した時**に心筋炎を発症する確率，

b) **接種しなかった時**に　〃

を比較すべきである．しかし，厚労省チラシは，上記 **b)** ではなく，

c) 接種しなかった時に，Covid-19に感染し「**入院**」した人が心筋炎を発症する「**条件付確率**」

と上記 **a)** を比較している．

(2) また，グラフ右側（及びデータ表最下行）の数値 $834/10^6$ は，上記 **c)**（**入院** した人の心筋炎発症率：$4/4798 = 834/10^6$）であるにも関わらず，

d) Covid-19に「**感染**」した人が心筋炎を発症する確率

としか読めない書き方をしている（*i.e.*,「**感染**」事象と「**入院**」事象を混同）．

確率空間

おまけのおまけ：念のための解説（続き）

(3) さらに，グラフ左側は **10代**，**20代** の（ワクチン接種後の）心筋炎発症率であるのに，グラフ右側の「国内」は，何故か，**30代を含めた年齢区分**（*i.e.*,接種する/しないの比較のみならず「海外」との比較からも不適切な年齢区分）となっている．

(4) グラフ左側の（ワクチン接種後の）心筋炎発症率は，国際的に信頼できる医学論文報告等の数値（*e.g.*, スライド p.24 で引用している Science News では **1/3000～1/6000**（15–24歳男性），NEJM (*1) 掲載論文では **1/6600**，最近のカナダからの報告では 約**1/5000**；いずれも Pfizerワクチン）と比べて「**桁違いに過小評価** (*2)」である．

(*1) NEJM (New England Journal of Medicine) は医学分野のトップ・ジャーナル．

(*2) 日本ではワクチン副反応の報告・調査システムが全く整備されていない（⇒ 医者が報告する義務もインセンティブもない）ため，ワクチン接種によって**実際に発生した副反応のごく一部（心筋炎の例では 約3%）しか厚労省に報告されていない**ことを（暗黙裡に）意味しています．