

遺伝学と統計学の歴史

スタージェン会長

鎌谷直之

Peter S. Harper

University Research Professor in Human Genetics

Cardiff University

Emeritus Professor of Medical Genetics

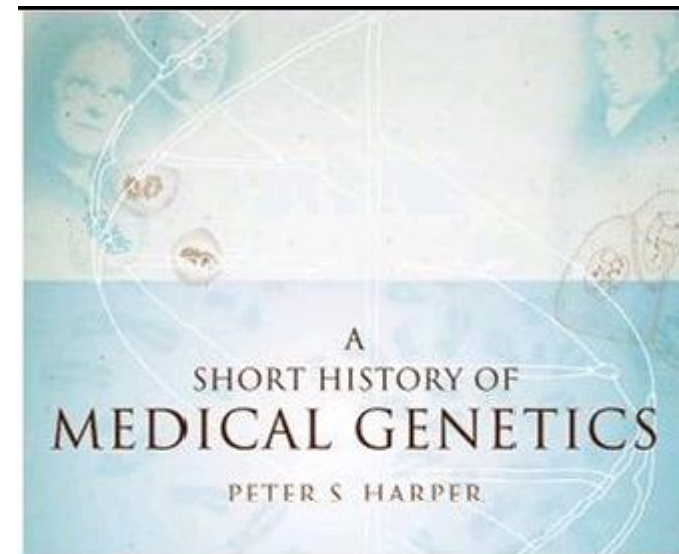
University of Wales College of Medicine

Cardiff, United Kingdom

OXFORD

UNIVERSITY PRESS

2008



Japan provides an unusual situation, for medical and human genetics have here been particularly weak, despite highly developed scientific, technological, and medical traditions. Mendelian genetics was taken up very early in Japan for the purpose of plant breeding (Matsubara, 2004), while

日本では遺伝医学と人類遺伝学が特に弱い!!!

strongly cultural isolation and extreme sensitivity over family matters, including genetic disorders, may have been delaying factors for medical genetics, as may the fact that much research on genetic disorders has been channeled through other medical specialties. A thorough study of all these aspects is needed. A paper on human genetics in Japan by Matsunaga (1992), while providing little information on medical genetics, indicates the persistence of eugenic practices and legislation.

It might be considered that religion would have been a major factor in

ベートソンが最初にgeneticsの用語を提案した手紙の下書き(1905年)

18.4.05.

TELEGRAPH, TELEPHONE
EXCH. HADLEY.
CRANFIELD, CAMBRIDGE.

Dear Ingham,

If the Quick Fund were used for
the foundation of a Professorship relating
to ^{Heredity & Variation} ~~Genetics~~, the best title
would, I think, be "The Quick Professor-
ship of the Study of Heredity." ~~Hereditarian~~!
No single word in common use ~~which~~
quite gives this meaning. Such a word
is badly wanted, and ~~it is more desirable~~ ^{if it were desirable} to introduce it. I
suggest "Genetics". Either ^{might be expressed} ~~clearly~~ clearly

If the Quick Fund were used for the foundation of a Professorship relating to Heredity and Variation, the best title would, I think, be "The Quick Professorship of the Study of Heredity." No single word in common use quite gives this meaning. Such a word is badly wanted and if it were desirable to coin one, "Genetics" might do. Either expression clearly includes Variation and the cognate phenomena.

遺伝学は「遺伝と多様性」の科学

遺伝学は

遺伝の科学

×

遺伝と多様性の科学

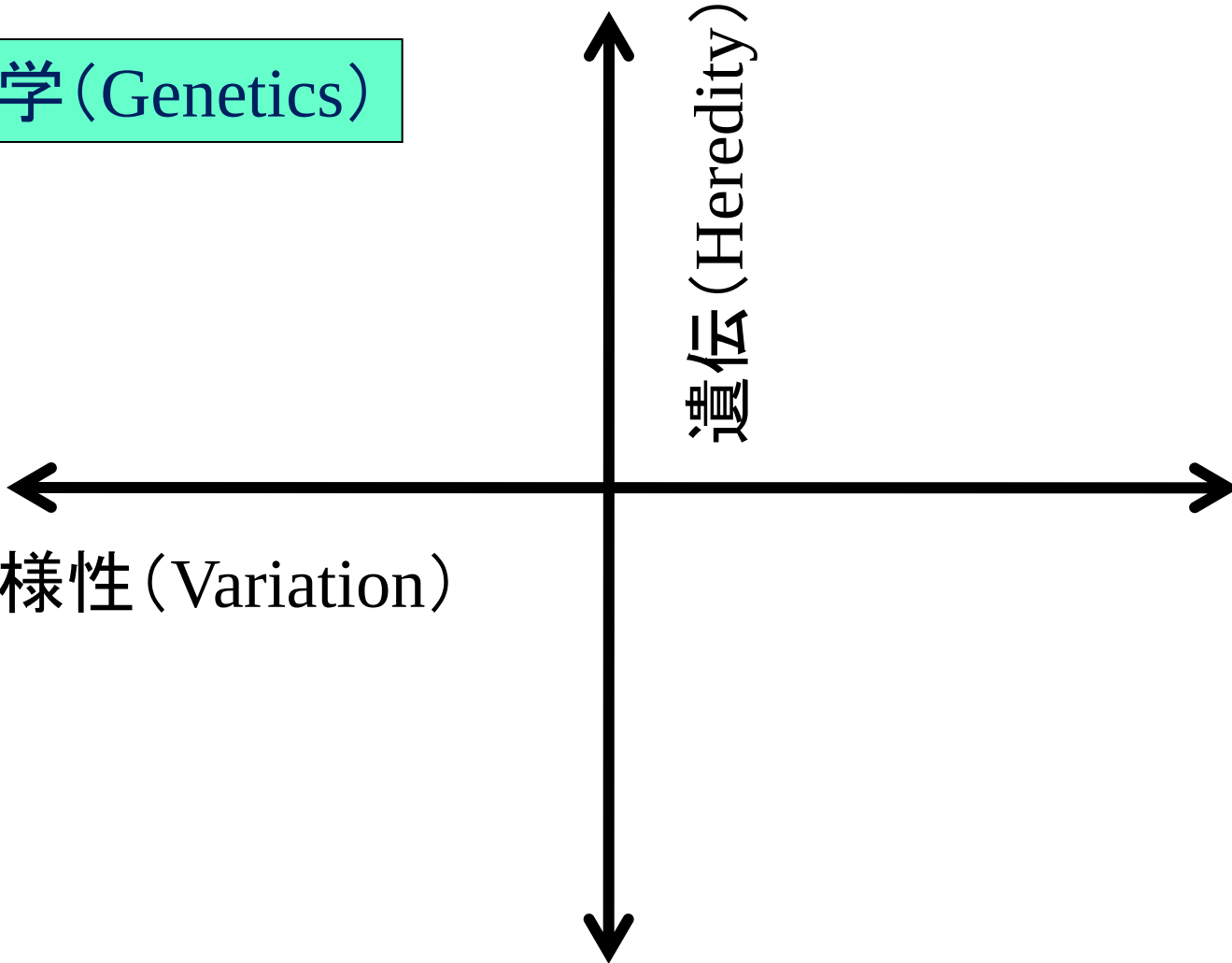
○

遺伝学は「遺伝」と「多様性」の科学

遺伝学 (Genetics)

遺伝 (Heredity)

多様性 (Variation)

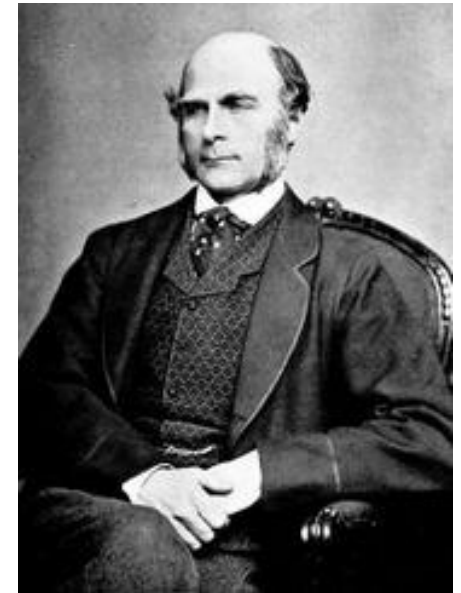
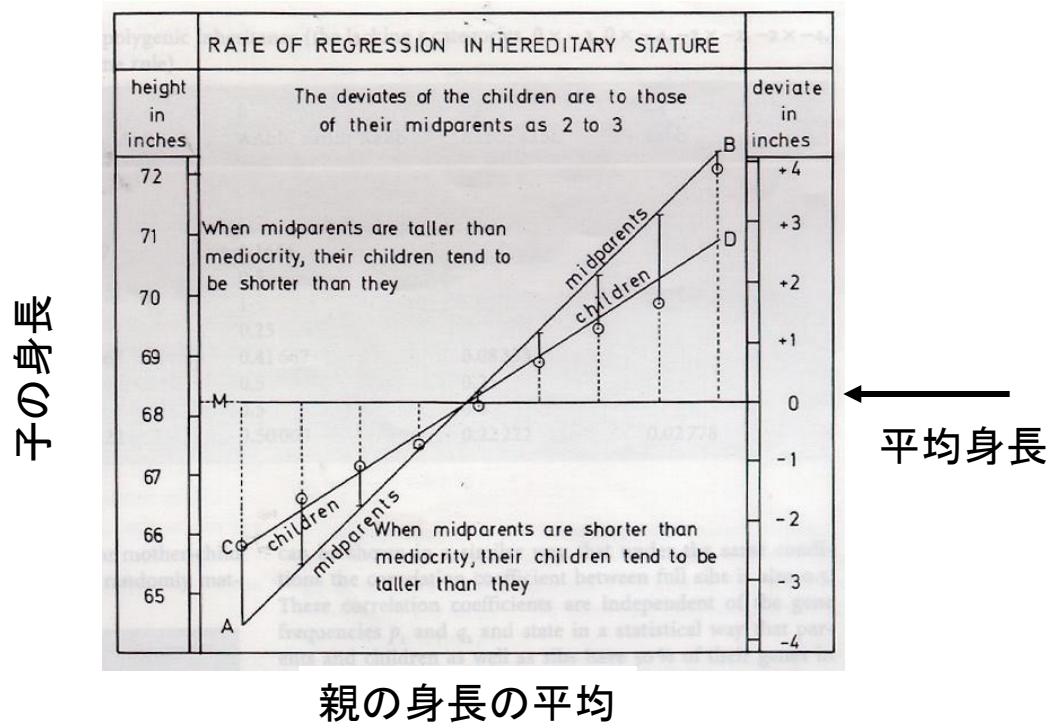


生物計測学派とメンデル学派の対立

19世紀前半まで、生物学は、リンネによる分類学、
ブルダッハなどによる形態学、ダーウィンによる
進化学が主流であった

数学は医学、生物学に応用されていなかった

ゴールトン(ロンドン大学)は回帰や相関など、基本的な概念を整備した



Francis Galton (February 16, 1822 – January 17, 1911),

$$E(\text{子の身長}) = a \times \text{両親の平均身長} + b$$

子の身長の両親の平均身長の上への回帰(Regression)を示すゴールトンの図

親の身長から子の身長を予測したい

ゴールトン

↓ 弟子

カール・ピアソン

この頃ロンドン大学に留学
したのが夏目漱石、日本で
始めて「遺伝学」のことばを
用いた



カール・ピアソンはロンドン大学で初代ゴールトン教授になった

カール・ピアソンは分布に関する数々の 概念を整備した

1. ピアソンの相関係数

$$\frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2} \sqrt{\sum_i^n (y_i - \bar{y})^2}}$$

2. ピアソンの χ 二乗検定

$$z^2 = \frac{(a + b + c + d)(bc - ad)^2}{(a + b)(a + c)(c + d)(d + b)}$$

3. モーメント生成関数やモーメントの概念

4. 異なった分布の間との関係と確率密度関数、分布関数

5. Biometrika, Ann Hum Genetの創刊者の一人



数学に真実を求め、現実世界にあてはめようとした
現実世界のデータをもとに、真実を発見する(帰納的手法)

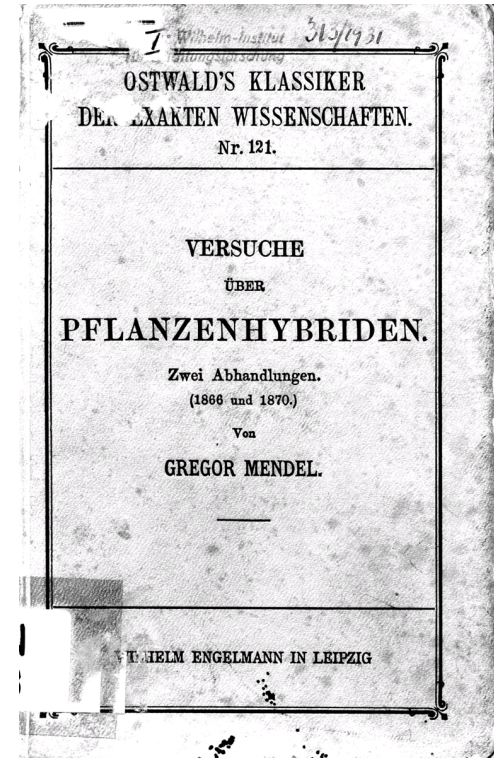
Karl Pearson (March 27, 1857 – April 27, 1936)

マルクス主義者

約100年前(1900年)、メンデルの法則が突然 再発見され、急速に生物学の進歩が始まった



1901年、ドブリスの変異
1902年、サットンの染色体説
1902年、ガローのアルカプトン
尿症の遺伝形式
1905年、ベートソン他、連鎖



A handwritten signature of Gregor Mendel in cursive script, reading 'Gregor Mendel'.

Gregor Johann Mendel (July 20[1],
1822 – January 6, 1884)

Gregor Mendel in der Zeit seiner Tätigkeit
als Forscher und Lehrer (um 1862).

生物計測学派 (Biometric school) とメンデル学派 (Mendel's school) の第一回目の衝突が起きた

帰納的手法

演繹的手法

連続

K ピアソン



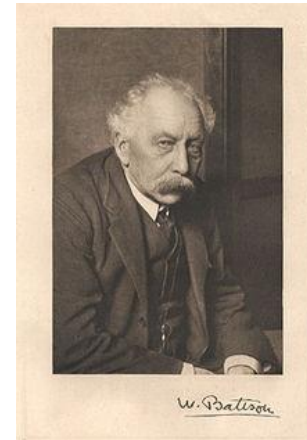
生物計測学派

弟子

メンデル学派

離散

ベートソン
(genetics、連鎖)



ヨハンソン
(gene、純系説)



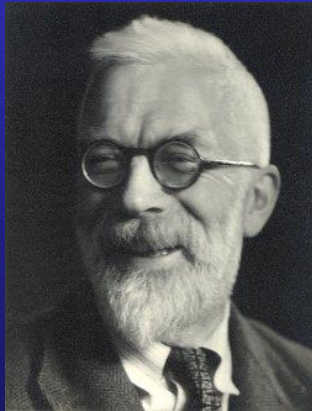
ウェルドン



現世界に真なる法則が存在しない

現世界に真なる法則が存在する

ピアソン



対立

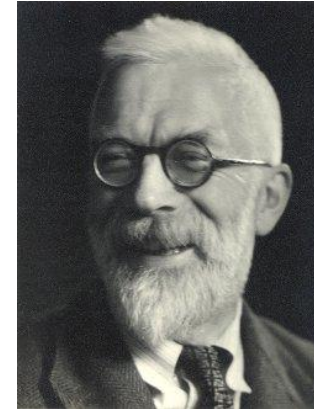
フィッシャー



フィッシャーはロンドン大学で二代目ゴールトン教授になった

生物計測学派(Biometric school)とメンデル学派(Mendel's school) の第二回目の衝突が起き、Modern synthesisが成立した

K ピアソン



フィッシャー

現世界に真なる法則が存在しない



現世界に真なる法則が存在する

-The Correlation between Relatives on the Supposition of Mendelian Inheritance. By R. A. Fisher, B.A. *Communicated by Professor J. ARTHUR THOMSON.* (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

Transactions of the Royal Society of Edinburgh, 52: 399-433 (1918)

Modern synthesisの成立と、近代統計学の成立を象徴する大論文
この論文が遺伝学と統計学の将来を決めた

XV.—**The Correlation between Relatives on the Supposition of Mendelian Inheritance.** By **R. A. Fisher**, B.A. *Communicated by* Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations σ_1 and σ_2 , it is found that the distribution, when both causes act together, has a standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$. It is therefore desirable in analysing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quantity the **Variance** of the normal population to which it refers, and we may now ascribe to the constituent causes fractions or percentages of the total variance which they together produce.

$$Var(X) = Var(G) + Var(E)$$

全分散 = 遺伝的分散 + 環境分散

多様性の尺度として、分散(Variance)という概念を提出した
GとEが独立なら、Xの分散はそれぞれの分散の和

$Var(G)/Var(X)$ は**遺伝力** (heritability)

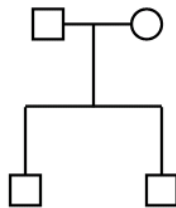
Gの関与は分散の比でわかる(分散分析)

XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. By R. A. Fisher, B.A. Communicated by Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

The values of the correlations between the representative measurements for random mating, which may be called the genetic correlations, are given in the accompanying table:—

Generations.	Half 2nd Cousin.	Half 1st Cousin.	Half Brother.	Ancestral Line.	Brother.	1st Cousin.	2nd Cousin.
Own	$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{32}$
Father's	$\frac{1}{128}$	$\frac{1}{32}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{64}$
Grandfather's	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{128}$
Great-grandfather's	$\frac{1}{512}$	$\frac{1}{128}$	$\frac{1}{32}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{256}$
Great-great-grandfather's	$\frac{1}{1024}$	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{128}$	$\frac{1}{512}$



メンデルの法則が正しければ、「遺伝型値」の
家族間の共分散や相関係数はexactに書ける

確率が正確に書ける事が
ゲノム情報の特徴(後述)

共分散や相関係数はデータから推定するもの、
勝手に与えるべきではない(ピアソン)

Modern synthesisへのフィッシャー(第二代ゴールトン教授)のその他の功績

線形モデルの始まり: 表現型に影響を及ぼす量的表現型は互いに相加的

$$X = \sum_i^n G_i + E$$

相加的ポリジーンモデル(離散と連続の問題を統合)

表現型値が遺伝型値(G_i)と環境値(E)の和なら
表現型値は正規分布に従う(中心極限定理)

仮説検定理論の導入 → 唯一の真なる仮説を検定する
ランダム化の重要性
フィッシャーの正確検定

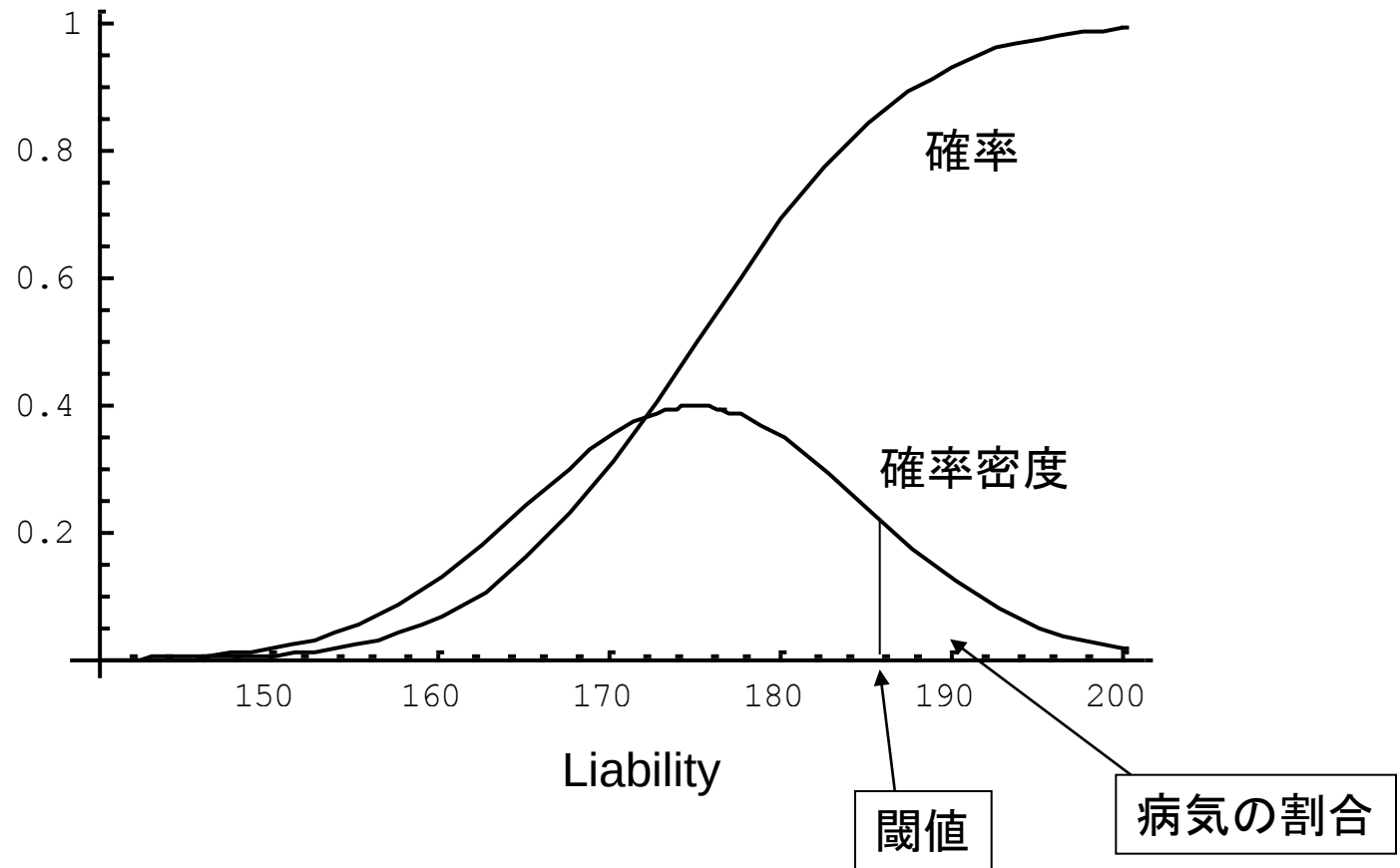
検定

尤度(likelihood)の概念を提唱し、最尤法による検定、推定の理論を完成した(観察データの得られる確率はexactに書ける) → 連鎖解析

推定

フィッシャー情報量

量的表現型と質的表現型の違い



量的変数はLiabilityを与え、Liabilityが閾値以下(以上)の人が病気になる

$$X = \sum_{i=1}^n G_i + E$$

Fisherの統計学と遺伝学との対応

統計学	遺伝学
1. 線形モデル	ポリジーンモデル
2. 分散分析	遺伝型値、環境値、遺伝力
3. 正確確率	メンデルの遺伝継承法則
4. 帰無仮説のみの仮説検定	メンデルの遺伝継承法則
5. 尤度、最尤法	連鎖解析

Fisherは農場に就職する前は遺伝学を主とした研究対象としていた

総合説の成立と生物学のパラダイム

Modern synthesis (総合説) の成立 (1920 - 1930年頃)

1. ダーウィンの進化説
2. メンデルの遺伝継承法則
3. 変異 (mutation)
4. 個体間の多様性 (variation): 生物計測学

フィッシャーはこれらを数理的に統合した

これ以降、生物学のパラダイムは変化していない
(数理生物学は現代生物学の成立に大きな役割を果たした)

遺伝統計学と数理統計学

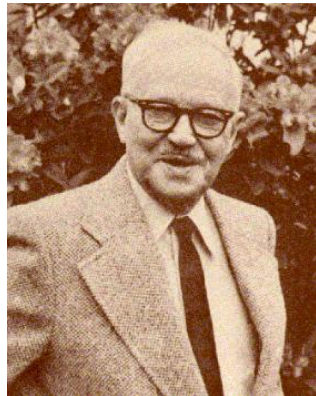
FisherはPearsonの息子とNeymanとまで対立した

Karl Pearsonの息子

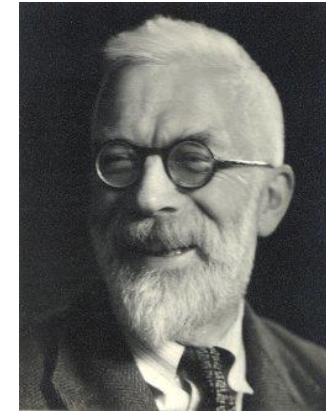


Egon Pearson +

Jerzy Neyman



対立



検定には帰無仮説と対立仮説が必要
ネイマン-ピアソンの補助定理
最強力検定

検定

真なる一つの仮説を設け検定

信頼区間

推定

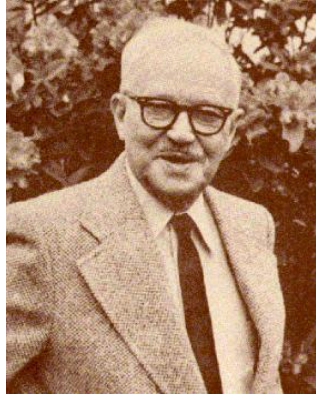
Fiducial interval

現世界に真なる法則が存在しない

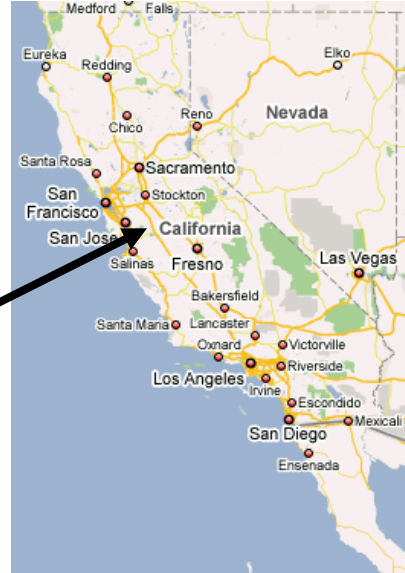


現世界に真なる法則が存在する

Neymanは米国に渡り、統計学部を作った



Neyman 1938
Berkeley



英国で開始され発展した統計学
が米国にわたった

Neymanはボレルやルベーグの教えを受けた
数学者、遺伝は知らない
遺伝以外の分野に統計学を応用するためには
FisherよりNeymanの方が良かった

NeymanはFisherに強い影響を
受けたが大嫌いだから後輩に
Fisherの功績を伝えていない

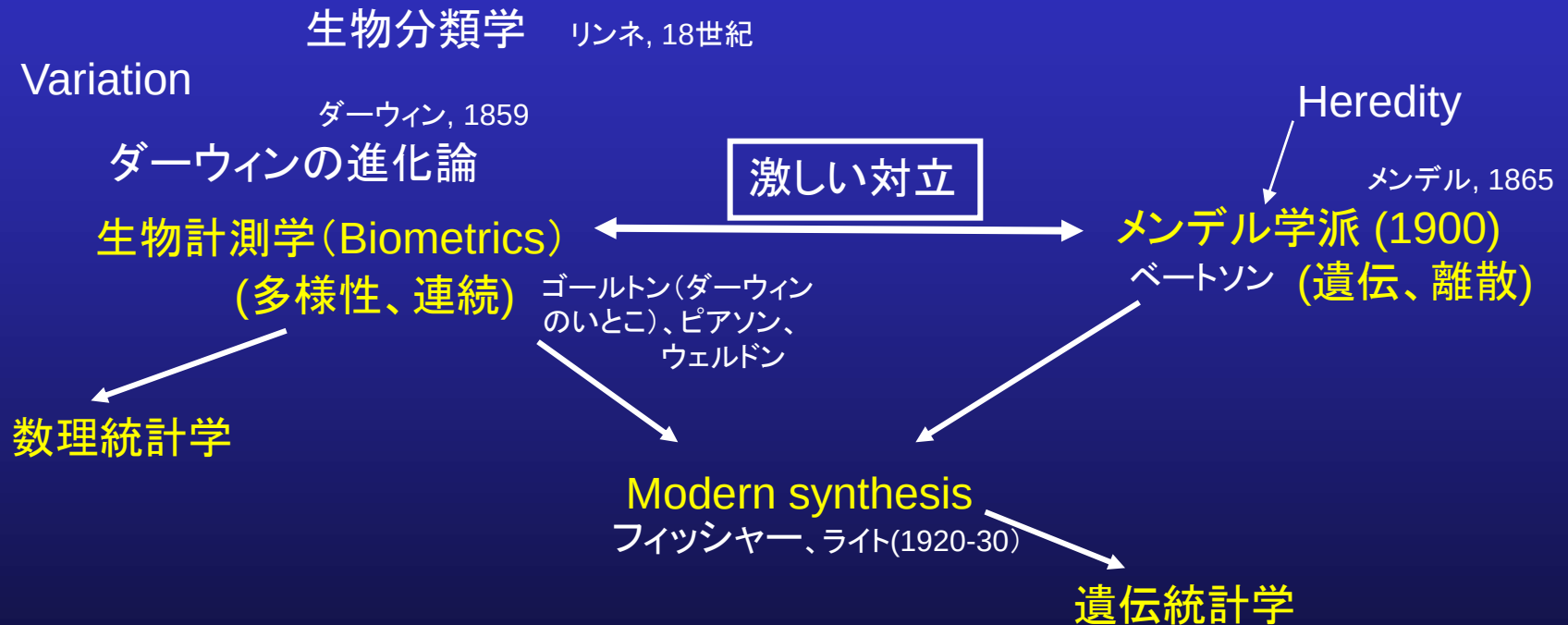
遺伝統計学と数理統計学の歴史

(まとめ)

1. 1900年以前は現実世界の真なる法則を仮定せずに、数理的に現実データ(遺伝的データ)を解析していた(ゴールトン、ピアソン)。
2. メンデルの法則の登場とともに、真なる法則が存在するという前提の下に遺伝のデータを解析する研究が始まった(フィッシャー)。
3. 現代生物学の基礎を作った総合説の成立には数理生物学が大きな役割を果たした。
4. しかし、遺伝以外の分野では、真なる法則が存在しないという前提の方が都合が良い(ネイマン)。
5. 遺伝統計学と数理統計学は以上の対立の歴史の上になりたっている。
6. このような対立の歴史を知らないと遺伝学と統計学の本質は理解困難。

統計学と遺伝学は共通の歴史を持つ

(統計学には二つの大きな流れがある)



数理統計学と遺伝統計学は考え方がかなり違う

なぜ、日本では遺伝学(特に人類遺伝学)が弱いのか

1. 遺伝学の定義が違う:

欧米では

「Genetics」は「HeredityとVariation(遺伝と多様性)」の科学

と定義されているのに、日本では

「遺伝学」は「遺伝(Heredity)」のみの科学、と考えられている

2. 日本の遺伝学のカバーする範囲が狭く、「広がりと深み」に欠ける

日本における遺伝学の系譜

なぜ日本にはvariationの概念が受け入れられ難いか

1. 日本語に可算名詞、不可算名詞の区別が無い
2. 日本語に単数と複数の区別が無い
3. 集合名詞の概念が無い
4. 島国のため、容貌などが異なる他の人種に接する機会が少なかった

日本における遺伝学の系譜

外山亀太郎

(写真: 竹内長正編, 外山直養

『外山亀太郎記念録』(1940年), 口絵より)

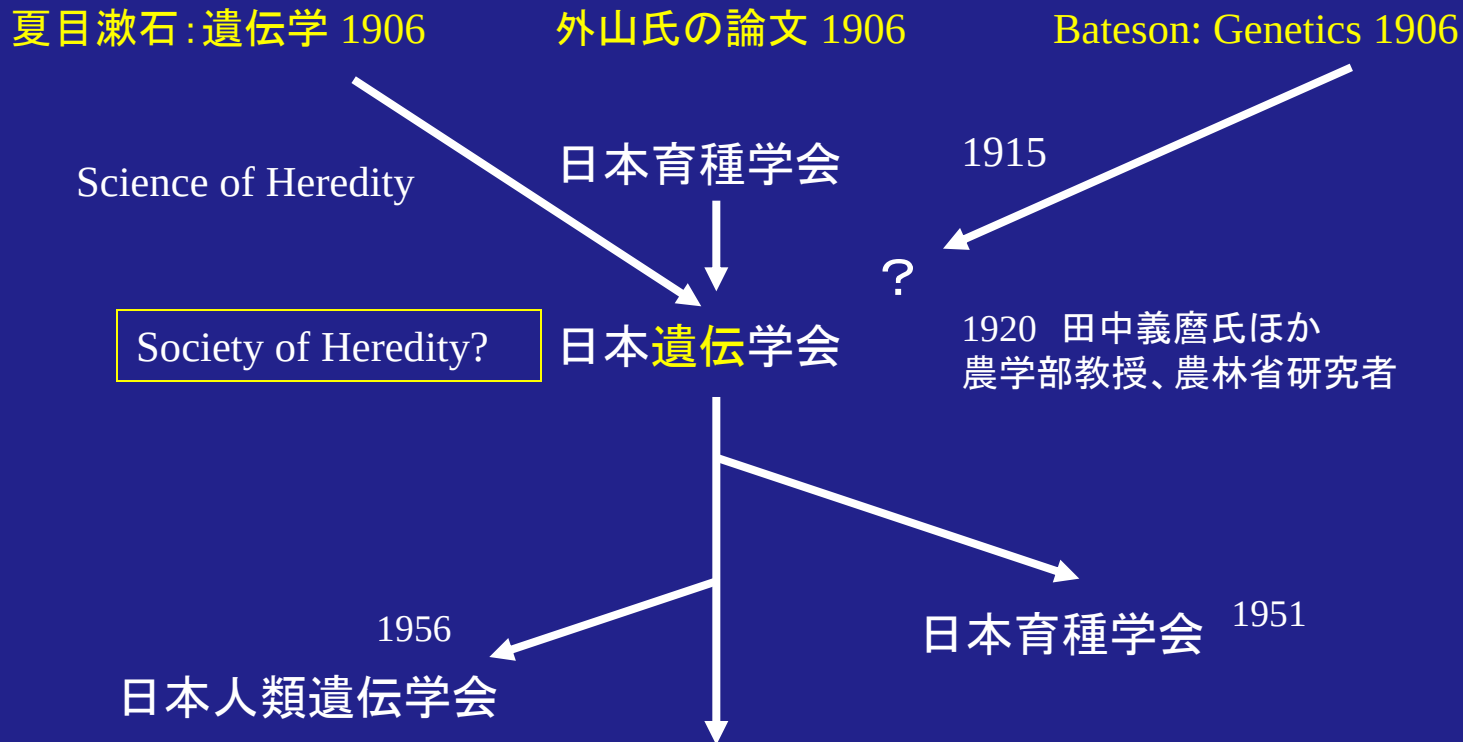


1906年、カイコの遺伝がメンデルの法則に従うことを明らかにした論文を発表

「昆虫の雑種学研究1、家蚕の雑種に就いて、特にメンデル遺伝法則を論ず」
英文、東京帝国大学農科大学学術報告第7巻第2号259-393, 1906

日本の遺伝学は「日本育種学会」として始まった

「遺伝学」の概念と日本遺伝学会設立の経緯



日本の遺伝学は育種と品種改良を目的に始まった(人間は対象外)
(この時、多様性は重視されず、Society of Heredityで問題なかった)

家畜や植物を対象とした概念や用語を、そのまま「人」に適用するには問題がある

初期の遺伝学の教科書におけるvariationの記述

1. 遺伝 Heredity

2. 変異 Variation ; Variabilität.

変異とは類似の反対で、生物の性質が互に相異なるを謂ふ。地球上同種の生物は無数に存するが、凡ての點に於て同一の個體と謂ふものは、古今東西を通じて一つもなく、無性生殖によつて生じた分枝群の中でも、瓜二つと云はれる双生児の間でも、必ず何處かに違つた點がある。「自然界に於て最も不變的な事實は變異の存在である」¹⁾と謂はれるやうに、變異ほど普遍的な現象はない。變異を分つて二種とする。

a) 彷徨變異 個體變異、連續變異、漸變、小變、不遺傳性變異、後天的變異。 fluctuation, continuous variation, individual v.; Modifikation, nichterbliche Veränderung, äußerliche V., Paravariation. その名の示すが如く、環境の影響により後天的に生じた變異で、遺傳性を有しない。環境と謂ふのは發芽、孵化または出産後のそればかりでなく、その以前でも受精後の影響は皆之に屬する。故にゴールトン F. GALTON が「天賦」nature の中に胎兒の母胎内に於て受けた變化まで含めてゐるのは賛成し難い。

b) 突然變異 偶然變異、不連續變異、急變、突變、激變、遺傳性變異、先天的變異。 mutation, discontinuous variation, genetic v.; Mutation, erbliche Veränderung, Heterogenesis, Idiovariation. 凡て遺傳性を有する變化を謂ふ。

彷徨變異と突然變異とは同じく變異の名を有するものの、その性質は殆ど根本的に異なつてゐると謂へる。彷徨變異は生物統計學的に研究されるものであり、突然變異は因子構造上から研究されるものである。彷徨變異を表す各個體を變員 variant; Variant と謂ひ、突然變異として出現した個體を突然變異型または突然變異者 mutant; Mutant と名ける。

1. 遺伝 Heredity

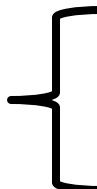
- | | | |
|-----------------|---|---------|
| 2. 變異 Variation | { | a. 彷徨變異 |
| | | b. 突然變異 |

この分類と訳が後々まで悪影響を与えている

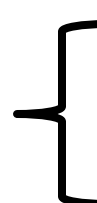
田中義麿:遺伝学
昭和9年初版、昭和25年第7版

正しい用語と分類

旧分類

- 1. Heredity (遺伝)
- 2. Variation (多様性) 
 - a. 彷徨変異 → variant
 - b. 突然変異 → mutant

正しい分類

- 1. Heredity (遺伝)
 - 2. Variation (多様性) 
 - a. 非遺伝的多様性
 - b. 遺伝的多様性
- variant

旧分類では、遺伝的変異 = 突然変異、とし、
多様性は環境によるもののみと考えている

状態としての「変異」と「遺伝的多様性」の関係

Mutation (変異)

Mutant (変異体)

Variation (多様性)

Variant (多様体)

頻度 (低)

起源 (新)

遺伝病

離散

(高)

(古)

健康 (多型)

連続



中間では、変異体と呼ぶか多様体と呼ぶかわからない。

連続的な概念だから当然!

遺伝学用語の定義の順番

用語	旧訳	作成者	年
Dominant	優性	Mendel	1865
Recessive	劣性	Mendel	1865
Allelomorph (Allele)	対立遺伝子	Bateson (Johannsen)	1902
Genotype	遺伝子型	Johannsen	1903
Mutation	突然変異	de Vries	1904
Genetics	遺伝学	Bateson	1905
Gene	遺伝子	Johannsen	1909

Geneの定義はgenotype, allele, geneticsの後

遺伝学用語改訂(日本人類遺伝学会)

英語	日本語	これまで
•genetics	遺伝学「意味: 遺伝と多様性の科学」	遺伝学「意味: 遺伝の科学」
•variation	多様性(バリエーション)	変異(彷徨変異)
•mutation	変異(突然変異)	突然変異
•variant	多様体(バリエーション)	変異体
•mutant	変異体(突然変異体)	突然変異体
•locus	座位	遺伝子座
•allele	アレル(アリル、アリアル)	対立遺伝子
•genotype	遺伝型	遺伝子型

()内は、許容される用語