

ベイズ統計学の原理と応用（その2）

美 添 泰 人

1 はじめに

本稿では、（その1）に続いて John W. Pratt の比較的初期の論文 [52] の内容を紹介する。その後で、ベイズ統計学を論理的に再構成した歴史的な著作である Leonard Jimmy Savage, *Foundations of Statistics* [57] のうち主観確率（ただし Savage はそこでは個人確率 *personal probability* と呼んでいる）を扱った2章および3章の主要部分を詳細な解説と証明を加えながら紹介する。

[57] は特に主観的な視点を公理的、体系的に導入したもので、統計学全体に極めて強い影響を与えたものである。竹内 [2] は、この本の出版を「第42章 統計学の復活」という表題で紹介し、それまで、ごく少数の例外を除いて否定的な扱いを受けていたベイズ統計学の転機となったものと評価している。Savage の提案によって、ベイズ統計学にも厳密な体系があり、非科学的と指摘されることがあった主観確率（個人確率）を用いない限り、統計的な判断、およびさらに一般的にデータに基づく意思決定は不可能であることが示されたといえる。その影響の一部は1959年に Savage が London を訪問した機会に Birkbeck and Imperial Colleges によって開催された Joint Statistics Seminar の記録 [59] で知ることができる。そこでは冒頭に Savage による講演が行われ、事前に指名された5名（M. S. Bartlett, G. A. Barnard, D. R. Cox, E.S. Pearson, C. A. B. Smith）が準備した講演が続き、その後で D.P. Armitage, I.J. Good, D. V. Lindley, H. Rubin などによって展開された活発な議論が記録されている。ここでは多くの議論のうち、二つだけ記したい。一つは I.J. Good の指摘とそれに対する Savage の回答

([59] p. 68–69) である。

Dr I. J. Good: It seems to me that one set of probabilities are tautological probabilities. They are probabilities whose values are assigned by *definition* of a statistical hypothesis.

Savage: Of course, many discussions and calculations in which we use the word probability are pure mathematics — they are schematic only — even when they are done with a view to immediate application. But it is anything but tautological to say that the errors made by a particular balance are normally distributed around the true weights with an unknown variance. It takes a lot of judgement to make such a statement responsibly, and we all know that this judgement can easily be wrong.

もう一つは、無作為化に関する D. R. Cox の指摘に対する Savage の回答 ([59] p. 86–87) である。ここで、ベイジアンでは観測値が得られた後の事後分布を利用するため、母数と無関係な無作為化は論理的に反映されないことに注意が必要である。

D. R. Cox: I would like Professor Savage to elaborate on remarks he made in his paper about the difficulty of justifying randomization from a strict Bayesian point of view. Part of the solution here may lie in attaching a particularly high utility to experiments for which many people can assign a reasonable prior distribution. If one thinks solely of a particular experiment desiring to produce closest possible estimates of a particular difference, then it seems reasonable sometimes not to randomize. One may do what Professor Savage said, namely to think up every little bit of information available and put it all together, and do what seems most likely to produce a precise estimate. But such an experiment may have very little value to anyone else, because not being aware of all the particular technical details, it would not be at all clear that there is not a tremendous systematic error in the experiment. One important property of randomization is that it makes the data reasonably convincing to other people as well as to oneself. Of course this is only half the story; randomization may increase accuracy by removing unsuspected biases. This aspect is particularly important in large experiments where bias is more important than random

error.

Savage: I think you lay your finger on the objectives of randomization, to make the experiment useful to others and to guard against one's own subconscious. What remains delicate and tentative for me is to understand when, and to what extent, randomization really can accomplish these objectives.

My doubts were first crystallized in the summer of 1952 by Sir Ronald Fisher. 'What would you do,' I had asked, 'if, drawing a Latin square at random for an experiment, you happened to draw a Knut Vik square?' Sir Ronald said he thought he would draw again and that, ideally, a theory explicitly excluding regular squares should be developed. As I have learned since, other statisticians have had, and worked on, this same idea; . . . This illustrates once more that one need not be a Bayesian to arrive at criticisms to which the Bayesian is led systematically.

The possibility of accidentally drawing a Knut Vik square or accidentally putting just the junior rabbits into the control group and the senior ones into the experimental group illustrates a flaw in the usual reference-set argument that sees randomization as injecting 'objective', or gambling-device probabilities into the problem of inference. If the randomization and the experiment were so executed by an automaton that no one knew which Latin square had been drawn or which animals had been put in the control group, the argument would, I suppose, apply. But, in fact, this information is not, and ought not to be, kept from the experimenter. And he ought not, in principle, to withhold it from those to whom he communicates his results.

In practice, we may hope, if the experiment is rather large and so designed as to control the variables that (subjectively) look most important, then randomization will almost always lead to a layout that does not look excessively suspicious to any given observer. But this hope needs serious investigation. Perhaps randomization is even one of the most efficient ways to arrive at such widely acceptable layouts. (Such rumours as that artists can make more random-looking designs than random number generators can are a little disquieting to this suggestion.)

In any event, randomization does remove an important possibility of personal interference, for anyone who believes that the randomization did take place according to Hoyle.

Many statisticians agree that an analysis of an experiment ought not be chosen at random. We think it wrong, for example, to break ties at random or to try to escape from the Behrens-Fisher problem by artificially pairing observations. But it has been puzzling to understand why, if random choices can be advantageous in setting up an experiment, they cannot also be advantageous in its analysis. The discussion Dr Cox and I have been giving of randomization seems to lead to an answer to this question. In making an analysis, there is no need to resort to chance to find a compromise analysis that will nearly suit everybody, for each interested person can in principle make for himself the analysis he thinks best. Attempting a compromise can only lose some of the relevant knowledge won by the experiment. Nor can randomization defend against the dangers of subconscious or conscious bias present at the analysis stage.

The arguments against randomized analysis would not apply if the data were too extensive or complex to analyse thoroughly by the individuals concerned. In such a case study of the data might itself become an empirical study based on sampling. Monte Carlo methods might be used. Or one of many possible expensive analyses might be determined in part by randomization in the hope of nearly pleasing everyone. It seems to me that, whether one is a Bayesian or not, there is still a good deal to clarify about randomization.

若干注記すると、ラテン方格 (Latin square) は n 行 n 列の表に n 個の異なる記号を、各記号が各行および各列に 1 回だけ現れるように並べたものであり、これを実験の順序が効果に与える系統的な影響を排除しつつ、精度の高い推定を可能とすることを目的として R. A. Fisher が実験計画で導入したものである。Knut Vik square とは、これに加えて二つの対角線上にも各記号が 1 回ずつ現れる形であり、偶然、この計画が選ばれた場合には、系統的な影響が混入することが予想される。なお according to Hoyle は「規則に従う」という意味の慣用句だが、Hoyle はコントラクトブリッジの原型である whist に関する規則本の著者であり、意思決定のゲームにからめて使いたくなった表現と思われる。

本稿では [59] についてはほとんど触れることができないが、別な機会に取り上げたい内容である。

2 Pratt の議論一つづき

前稿「ベイズ統計学の原理と応用（その 1）」[14] では、Pratt [52] のうち、§1. Introduction から、§2. Insufficient statistics, §3. Estimation, §4. Confidence regions までを紹介した。本稿では、§5. Maximum likelihood 以下の内容について紹介する。前稿に準じて、以下では小節の番号および表題を [52] の番号に合わせて明示する。

2.1 最尤推定

最尤推定に関しては従来から多くの議論があり、Pratt [52] には特に新しい内容は含まれていないが、話題の重要性を考慮して、いくつかの点に関する議論を要約している。未知の（一般には多次元の）母数を θ とするとき、その最尤推定量 $\hat{\theta}$ は、事前分布の密度が $\hat{\theta}$ の近くで一定、かつ離れた領域ではそれほど大きくない場合は事後モードに一致する。また、大標本では、通常、事後期待値に近い。

したがって、Pratt が信頼分布（confidence distribution）とよぶ分位点の分布は事後分布で近似され、 $J^{1/2}(\hat{\theta} - \theta)$ が漸近的に正規分布にしたがうという性質から導出される。 $J^{1/2}$ は、 $J^{1/2}J^{1/2'} = J$ を満たす任意の正方行列であり、 J としては Fisher 情報行列 $J(\theta) = E \left[\left(\frac{\partial \log f}{\partial \theta} \right) \left(\frac{\partial \log f}{\partial \theta} \right)' \mid \theta \right]$ 、または $J(\theta)$ の最尤推定量 $J(\hat{\theta})$ 、または標本情報行列 $\tilde{J}(\hat{\theta}) = \left(\frac{\partial \log f}{\partial \theta} \right) \left(\frac{\partial \log f}{\partial \theta} \right)'$ が利用される。ここで f は確率変数の密度関数を表す。伝統的な立場からは $J(\theta)$ を用いることが自然であるが、 J が θ に依存しない場合、信頼領域 $\{\theta : J^{1/2}(\hat{\theta} - \theta) \in R\}$ の評価がはるかに容易となることから、 J の代わりに $J(\hat{\theta})$ が用いられることもある。さらに、もう一つの可能性として対数尤度の 1 階偏微分ベクトルが近似的に正規分布に従うことを用いることもある。このベクトルを $\hat{\theta}$ のまわりで 1 次式に展開して近似すると $J = \tilde{J}(\hat{\theta})J^{-1}(\hat{\theta})\tilde{J}(\hat{\theta})$ が得られる。ベイズアンの視点で対数尤度を $-(1/2)(\hat{\theta} - \theta)'J(\hat{\theta})(\hat{\theta} - \theta)$ で近似すると最も自然な選択は $J = \tilde{J}(\hat{\theta})$ である。このことは Jeffreys [37] のほか、Lindley や Savage も指摘していて、現在では広く知

られているように、 $\tilde{J}(\hat{\theta})$ と $J(\theta)$ の違いは得られた精度と期待される精度の違いである。Fisher [29] は、新しいデータと過去のデータを結合するために $\tilde{J}(\hat{\theta})$ を補助統計量 (ancillary statistics) として導入した。補助統計量は関心のある母数 (この場合は θ) に依存しない分布にしたがうもので、美添 [3, 13] でも触れているとおり、伝統的な統計学の推論では補助統計量の値は固定して考えてもよいという「条件性の原理」が支持されている。この原理に対する批判も、美添 [3, 13] で紹介している。Pratt は、どのような条件があれば、伝統的な立場の推論に際して $\tilde{J}(\hat{\theta})$ を固定した条件付の推論が正当化されるのか、またどのようにして正当化できるのかは、明確ではないと述べている。いずれにしても、ベ이지アンの立場からは J として $\tilde{J}(\hat{\theta})$ を選択する議論は論理的に明瞭である。

2.2 近似に関する注釈

この論文において Pratt は、標準的な推論手法をベ이지アンの枠組みからの近似として解釈してきた。これは一般的な近似と共通する側面を持ち、例えば、一般には複数の近似が利用可能であり、そのうちの一つが優れていると考える根拠があればともかく、そうでないときは全体的な意味でそれほど悪くない手法で満足することになる。

多くの応用例では、通常、近似の正確性に関する情報は多くないか存在しない。計算可能性という側面から誤差に関して極度に保守的な限界とならない近似で満足できる場合はあるが、そのような限界が得られない応用例は無数にある。

今考えている近似が多くの近似と違う点は、近似を正当化する根拠が弱いこと、実際には極端に弱いということである。ベ이지アンの視点からは、これが伝統的な手順に関して真に危惧すべき点であり、複数の手法が存在したり、それらの近似の精度が未知であるという問題とは異なると、Pratt は主張する。

美添 [14] で紹介したとおり、Pratt [52] の §3. Estimation と §4. Confidence regions では、「典型的な状況の場合」の近似を検討している。たとえば §4 では、観測値が与えられたときの事後分布における事象 ($\tau \in R$) の (条件付) 確率が、そ

の事象のすべての可能な観測値に関する周辺分布の確率に等しいことを要請する。ただしこの例では両方の確率は母数に関しては周辺確率となっている。正確に事後分布が評価できるのであればこのことは問題にはならないが、考えなければならないのは、計算上の困難や、事前情報を適切に事前分布として表現する難しさから、事後分布の評価ができない場合である。このときには、手元の観測値が特殊で、事象 ($\tau \in R$) の事後確率が、すべての可能な観測値に関する周辺分布と異なるのかどうかについて、非公式な判断が必要となる。「どのような状況が典型的か」という質問に明確に答えることは困難であるが、伝統的な手法においてもどのような情報に関する条件を所与とするかは、同様な難問である。このあたりの Pratt の説明は難解だが、例として順序統計量の間隔が広い場合に適用される適当な平滑化を上げている。

別な例として、事前分布が「散漫な分布」とは異なってある程度情報を含んでいる場合には近似は貧弱となると指摘する。推定値または信頼区間の位置が事前分布と比較して大きく異なるものであれば、平均または中央値の事後分布は伝統的な手法とは異なって事前分布の中心方向に向かうし、位置が大きく異なる場合でも伝統的な信頼分布は事後分布より幅が広がる。この議論は Pratt [50, 51] に記されているが、ここでは省略する。

さらに、Pratt は（強い）尤度原理に言及している。美添 [13] でも説明したとおり、「強い尤度原理」とは、同一の θ を母数とする異なった確率分布に従う確率変数 x, y について、それらの尤度関数に $p(x_1, \dots, x_m | \theta) \propto p(y_1, \dots, y_n | \theta)$ という比例関係がなりたつときは両者の推論が一致することを要求するものである。ベイズ統計では、 θ の事前分布が同一なら事後分布は一致するから、強い尤度原理が成立する。これに対して、「弱い尤度原理」は同じ確率分布 $p(x | \theta)$ にしたがう2組の観測値 $x_1 = \{x_{1i}\} (i = 1, \dots, n_1), x_2 = \{x_{2j}\} (j = 1, \dots, n_2)$ について、それらの尤度関数が比例的 $p(x_1 | \theta) \propto p(x_2 | \theta)$ であれば x_1 と x_2 は同じ推論を導くことを要求する。この原理は、最小十分統計量を用いる推論と同等である。

伝統的な統計学では、弱い尤度原理は認めるが、それより厳しい条件である

強い尤度原理は認めない。たとえば、不偏性は強い尤度原理を満たしていないことは二項分布と負の二項分布の例でもわかる。Pratt は、伝統的な統計学が強い尤度原理を認めないとしても、それが原理ではなく近似の問題であるなら大きな問題ではないとは言え、強い尤度原理を満たさない伝統的な手法を、強い尤度原理を満たすバイジアンの手法で近似することを意味する点が問題だと主張する。

以下 Pratt は強い尤度原理に関する解説を与えている。その内容はすでに美添 [13] で紹介しているので、ここでは省略するが、Pratt が指摘していない点があり、実は、伝統的な統計学では「条件性の原理」を認めている。これは、ある確率変数の分布が母数 θ に依存しないとき（これを補助統計量と呼ぶ）、 θ に関する推論では、補助統計量の値を固定した条件付き分布を用いることを意味するもので、広く受け入れられている。ところで、Birnbaum [20] は、弱い尤度原理と条件性の原理から強い尤度原理が導かれることを示している。[20] には、バイジアン立場から Savage と Pratt の興味深いコメントがあるが、残念ながらここで紹介する余裕がない。

2.3 仮説検定：有意水準

§7 で仮説検定を扱うが、まず、1 次元母数 θ に関して実数値を取る検定統計量 T を用いる例を考える。ここで攪乱母数の存在、すなわち θ 以外にも T の分布を特定する母数が必要となる場合を許容する。

帰無仮説を $\theta \leq 0$ 、対立仮説を $\theta > 0$ として T が大きいときに仮説を棄却するものとする。典型的には $\theta = 0$ のときに T の分布は確定し、その分布は $\theta < 0$ のときの T の分布より確率的に大きい (stochastically larger)。ここで、分布 F_1 が F_2 より確率的に大きいとは $F_1(t) < F_2(t)$ 、いいかえれば対応する確率について $P_1(T \leq t) > P_2(T \leq t)$ がなりたつことを表す。

この状況で $T = t$ が観測されたとき、この検定は $\alpha \geq P_0(T \geq t)$ を満たすすべての水準 α で仮説を棄却する。ただし $P_0(T \geq t)$ は $\theta = 0$ のときの $T \geq t$ となる確率を表す。したがって $T = t$ は水準 $P_0(T \geq t)$ でちょうど有意となる、すなわち

P 値は $P_0(T \geq t)$ である。信頼限界と検定の関係から、水準 $1-\alpha = 1 - P_0(T \geq t)$ における θ の信頼限界は $\theta = 0$ となる。このことは、一般に、水準を α とする検定統計量によって仮説 $\theta = \theta_0$ が受容されることが、 $P(T < t \mid \theta_0) \geq 1 - \alpha$ を意味することから確かめられる。

ここで信頼水準 $1-\alpha$ を事後確率の近似と解釈するなら、 P 値 $P_0(T \geq t)$ は $\theta < 0$ となる事後確率である。つまり、 $1-\alpha = 1 - P_0(T \geq t) \simeq P(\theta \geq 0 \mid t)$ から $\alpha \simeq P(\theta < 0 \mid t)$ となる。仮説検定における多くの状況のように、あらかじめ θ は 0 に近い、ないし $\theta < 0$ の可能性は低いと想定する場合、特にその判断が標本より強い場合は、この近似は悪くなると予想されるのは当然である。

ここで Pratt は θ が T の分布の translation parameter となる場合、すなわち分布関数が $F_\theta(t) = F_0(t-\theta)$ 、密度関数は $f_\theta(t) = f_0(t-\theta)$ 、と表される場合を取り上げる。このときは、ベ이지アン視点からは $T-\theta$ と θ は独立である。これは $u = t-\theta$ とおくと、 $du/dt = 1$ だから、密度関数で表せば $f_\theta(t) = f_0(t-\theta) = f_0(u)$ となり、右辺が θ に依存しないことから確かめられる。さらに θ の事前分布が散漫なら $T-\theta$ と T も独立となる。このことは、事前分布が $p(\theta) = \text{const}$ なら、 (t, θ) の同時分布は $p(t, \theta) = p(\theta)p(t \mid \theta) \propto f_0(t-\theta)$ となるが、 $t = u+\theta$ として変数を (t, θ) から (t, u) に変換すると (t, u) の同時分布は $p(t, u) = f_0(u+\theta-\theta) = f_0(u)$ となり、最後の表現が t を含まないことからわかる。したがって、 P -値は、 T が与えられたときの $\theta < 0$ となる事後確率と一致する。

以上の片側 P -値の解釈は、 T にもとづく信頼限界が定義されない状況でも適用できる。 $\theta \neq 0$ のときの T の分布が完全に確定できないときは、信頼限界を定義することには問題があるが、このような条件の下でも、水準 $1-\alpha$ の信頼下限は、 P -値より大きい（小さい） $1-\alpha$ に対して 0 より大きく（小さく）なるだろうから、 P -値は $\theta < 0$ となる事後確率を上記と同じ意味で近似している。たとえば、ある処理をブロックごとに対照と比較する場合には $\theta = 0$ は処理効果が存在しないことを表す。処理効果が存在する場合には各ブロックにおいて同じ符号を取ることを仮定するとき、どのような θ の定義や信頼区間の構成方法に対しても P -値に関する上述の解釈は成立する。

次に、帰無仮説を $\theta = 0$ 、対立仮説を $\theta \neq 0$ として、 T が大きいまたは小さいときに棄却する場合、具体的には水準 $\alpha/2$ の片側検定のどちらかで棄却する場合を想定する。片側 P -値に関する上述の解釈が成立する状況では、0 に関して標本の示唆する側とは反対側に θ が含まれる事後確率は、近似的に、両側検定の P -値の「半分」となる。これは観測値によってどちらの片側検定を利用するかを決定する方法で、伝統的な統計学の教義では完全に異端である。一方で、以下に記す信頼区間の手順は、これと同じことであるが、これは最も正統的な頻度論者でも認める手順であり、 α を 2 倍することなく、データに依存して方向が決まる命題となる。

- (a) T が水準 α の下方臨界値より小さいとき、 $\theta \leq 0$,
- (b) T が水準 α の上方臨界値より大きいとき、 $\theta > 0$,
- (c) それ以外のとき、 $-\infty < \theta < \infty$ と宣言する。

もし $\theta \leq 0$ であれば (b) だけが誤りで、その確率は α 以下である。またもし $\theta > 0$ であれば誤りを犯す確率は α 以下である。これは相互に排反な二つの帰無仮説について同時に水準 α の検定を実施する手順とみなせる。

両側検定の P -値と事後確率に関する自然な関心に対して、広範囲に適用可能な関係は、以上のとおりである。帰無仮説 $\theta = 0$ に対する両側検定の P -値は、多くの場合、 θ の推定値 $\hat{\theta}$ との距離が 0 と $\hat{\theta}$ との距離より大きい、すなわち θ が 0 と $2\hat{\theta}$ の間に入らない、事後確率の近似を与える。しかし、この特別な区間の確率を求める明確な理由は見当たらない。ここで Pratt は、Jeffreys [37], Lindley [45] の例を参照しながら、帰無仮説 $\theta = 0$ の確率が正であっても、「両側検定の P -値は $\theta = 0$ となる事後確率とは何ら特別な関係はない」と記している。その意味は次のとおりである。

$x_1, \dots, x_n$ を正規分布 $N(\theta, 1)$ からの無作為標本として、仮説 $H_0 : \theta = 0$ を対立仮説 $H_1 : \theta \neq 0$ に対して検定する問題で、有意水準を α とすれば、棄却域は $|\bar{x}| > c/\sqrt{n}$ で与えられる。ただし、 c は標準正規分布の上側 $100\alpha/2\%$ 点である。ベイジアンでは、Jeffreys [37] のように事前分布として $P\{\theta = 0\} = \pi$ と $\theta \neq 0$ の

ときの密度関数 $p(\theta)$ を想定することができる。ただし $\int p(\theta)d\theta = 1 - \pi$ （積分範囲は $\theta = 0$ を除く母数空間）とする。ここで標本平均 $\bar{x}$ が与えられたときの事後確率を計算すると

$$P\{\theta = 0|x\} = \frac{\pi \exp - \frac{n\bar{x}^2}{2}}{\pi \exp - \frac{n\bar{x}^2}{2} + \int \exp \left[-\frac{n}{2}(\bar{x} - \theta)^2 \right] p(\theta)d(\theta)}$$

となるが、この確率と両側検定の P -値とは無関係である。同様に、両側検定の P -値は、 θ が小さい（たとえば $|\theta| < \delta$ となる）事後確率とも無関係である。

なお、伝統的な統計学が強い尤度原理を認めない理由の一つに、この問題がある（美添 [3] 参照）。強い尤度原理によれば、どのような停止規則を用いても結論は変わらないが、標本平均 $\bar{x}$ がその限界 $c/\sqrt{n}$ を超えるまで抽出を続けると、実際に仮説 $H_0: \theta = 0$ が正しいときでも確率 1 でこの限界に到達するから、強い尤度原理を受け入れるなら、常に H_0 を棄却することを認めざるを得なくなる。さらに補足すると、 H_0 が棄却される値 $\bar{x} = c/\sqrt{n}$ を代入して $n \rightarrow \infty$ とすると、 $P\{\theta = 0|x\} \rightarrow 1$ となって、ベイズアン立場で考えれば、データは $\theta = 0$ であることをはっきり示している。

次に θ が k 次元 ($k > 1$) の場合を取り上げる。母数の半空間、または十分平坦な $k-1$ 次元超曲面によって、母数空間を近似的に二つに分割する仮説ならば、1 次元母数に関する片側検定と本質的に類似している。同様に、 θ が k 次元であり、帰無仮説が「 θ が $k-1$ 次元超平面または十分平坦な $k-1$ 次元超曲面の上にある」であれば、1 次元母数に関する両側検定と本質的に類似している。すなわち、 $k-1$ 次元超曲面に関して、 θ が観測値が示唆する範囲の反対側にある事後確率は、前者であれば P -値、後者であれば P -値の半分で近似される。実用上はこのような状況は多くないが、例として、攪乱母数と 1 次元母数に関する片側検定または両側検定として自然に表現される場合がある。たとえば、半平面の仮説 $\theta_1 \leq \theta_2$ は、 $\theta' = \theta_1 - \theta_2$ とおけば $\theta' \leq 0$ と表され、超平面の仮説 $\theta_1 = \theta_2$ は $\theta' = 0$ と表される。

通常の帰無仮説では θ は $k-1$ 次元より小さな次元の曲面の上にある。このような場合、 P -値はベイズアンにとってはまったく役に立たない。ただし、例

外として次の解釈がある。すなわち帰無仮説 $\theta = 0$ の検定における P -値は、「 $\hat{\theta}$ と θ との距離が $\hat{\theta}$ と 0 の距離より大きい事後確率」の近似となることが多い。残念ながら、ここで用いられる距離は奇妙なものになるであろう。なぜなら、この距離は検定統計量によって定められるが、通常、統計量は意味のある距離にもとづいて選ばれるのではなく、帰無仮説の下で便利のように選ばれるからである。

これまでに触れていない問題に、1 次元母数 θ が非負の場合の帰無仮説 $\theta = 0$ があり、例として決定係数 R^2 または非心母数 (non-central parameter) が 0 という仮説が考えられる。適当に書き直すと、これらの例は k 次元の母数 θ に関して、 θ が $k-1$ 次元より小さな次元の曲面の上にあるという帰無仮説として表現されるから、前の段落の指摘が該当して、 P -値は役に立たない。

結論として、1 次元母数 θ の母数空間が $\theta < 0$ の範囲を含み、帰無仮説 $\theta \leq 0$ を対立仮説 $\theta > \leq 0$ に対して検定する場合には、 P -値は通常は $\theta \leq 0$ となる事後確率を近似する。 P -値について有用な解釈が可能なほとんどの状況は、この形になっている。なお $\theta \leq 0$ は $\theta \leq \theta_0$ または $\theta \geq \theta_0$ で置き換えることができる。 P -値と事後確率の関係は分散または分散共分散行列が既知の場合の正規分布に関しては、容易に確認することができる。

ここで Good [32] が示した不等式に言及している。 H_0 と H_1 を互いに排反な仮説として、統計量 T が大きいときに H_1 を支持するものとする。このときには $T \geq t$ のときは $T = t$ のときより強く H_1 を支持するから、

$$\begin{aligned} \frac{P(H_0 | T = t)}{P(H_1 | T = t)} &> \frac{P(H_0 | T \geq t)}{P(H_1 | T \geq t)} = \frac{P(H_0 \cap T \geq t)}{P(H_1 \cap T \geq t)} \\ &= \frac{P(H_0)P(T \geq t | H_0)}{P(H_1)P(T \geq t | H_1)} > \frac{P(H_0)}{P(H_1)}P(T \geq t | H_0) \end{aligned} \quad (2.1)$$

が成立する。この式は、 $T = t$ のとき、 H_0 の H_1 に対するオッズは事前オッズと $P(T \geq t | H_0)$ の積より大きいことを示している。最後の因子は、 H_0 が $\theta = 0$ を意味するとともに T の分布を特定する場合には、 P -値である。残念ながら、応用例の大多数では検定の妥当性は疑わしいと Pratt は指摘している。

2.4 検定の一般的な利用について

Pratt [52] の §8 では、ベイジアン視点から、検定の一般的な利用の妥当性を検証している。なお、Edwards-Lindman-Savage [27] は、同じ視点からさらに豊富な例を扱っている。一般的な注意としては、仮説検定は行動が二つの場合の意思決定問題とみなすことができ、許容される検定はベイズ解のクラスとなる。以下では [52] が取り上げた話題の一つを紹介する。

二つの状態を単純仮説 H_0 と H_1 として、行動が二つの決定問題を考える。これは分類の問題と考えてもよいが、このような問題はいずれも帰無仮説 H_0 に関する検定とみなすことができ、 α は H_0 が真のときに好ましくない行動を選択する確率となる。 H_1 が真のときに好ましくない行動を選択する確率を β とすると、達成可能な (α, β) の点から一つを選ぶ決定問題となる。2 種類の誤りに対する損失が等しいとき、伝統的な枠組みでは許容される検定（今の例では最強力検定または尤度比）で $\alpha = \beta$ を選択することが示唆される。

これは、事前確率において H_0 と H_1 とが同様に確からしく、達成可能な (α, β) 曲線が α と β に関して対称となる場合のベイジアン手続きと一致する。

H_0 と H_1 が同様に確からしくても、 α と β に関する対称性がない場合は、ベイジアンでは α または β の微小な減少が他方における等しい増大となるような (α, β) 曲線上の点を選択される。その点では尤度比が 1 となるが、一般には $\alpha = \beta$ とはならない。事前確率および損失が等しくない場合には、ベイジアンの手順は $k\alpha + \beta$ を最小にするように選ばれる。ここで k は事前オッズと損失の比率の積である。Pratt はこれ以上の説明をしていないので、その導出を紹介する。

H_0 が正しいときに H_1 を選択する損失を $L_0 = L(H_1; H_0)$ 、 H_1 が正しいときに H_0 を選択する損失を $L_1 = L(H_0; H_1)$ と表し、さらに $\alpha = P(H_1 | H_0)$ 、 $\beta = P(H_0 | H_1)$ と書くと、リスクは $R = P(H_0)P(H_1 | H_0)L_0 + P(H_1)P(H_0 | H_1)L_1$ だから、 $R = P(H_1)L_1(k\alpha + \beta)$ と変形できる。ただし $k = P(H_0)L_0 / P(H_1)L_1$ である。 $P(H_1)L_1$ には (α, β) は含まれないから、 $k\alpha + \beta$ を最小にすればよい。ところで、この点の傾き、すなわち臨界点の尤度比は $-k$ である。このように、ベイジアンでは α と β の交換に関する比率、ひいては尤度比そのものを考慮する

のに対して、伝統的な手法ではその確率を基準としている。 (α, β) については、Lindley [47] には無差別曲線が平行な直線になるという興味深い解説があるが、ここでは省略する。

最後の §9. Miscellaneous remarks において Pratt は次のように要約している。ベイズ統計学の視点から伝統的な手法を正当化する根拠は強いとはいえないが、それはこれらの手法の背後に避けられない欠点があるためである。ベイジアンと伝統的な手法は、いずれも与えられたデータの状況をそれが発生した状況を通じた平均によって正当化するという類似点があるが、ベイジアンでは事前分布による平均、伝統的な統計学では無限の（仮想的な）観測値による平均を用いる点で異なっている。

3 Savage による主観確率の公理体系

以下では、ベイズ統計学の基礎を与えた Savage [57] が展開した、主観確率の公理体系を紹介する。そこで Savage は個人確率（personal probability）と呼んでいるので、以下でもその用語を用いる。なお参照のために、原書における章、節の番号を §2.2 などと明示する。また定理などの番号は節ごとに改められて引用の際に節の番号が明示されるが、本稿では例えば 3 章 2 節の定理 1 は「定理 3.2.1」と記す。

3.1 不確実性の下での決定問題

3.1.1 意思決定者

§2.2 The person で、Savage は意思決定に関して合理的な行動をとる高度に理想化された個人（The person）を想定する。その際、どのような行動規範が「合理的」であるかについて合意が必要であるが、合理的が論理的を意味する限り、それ以上の問題はない。ただし、脚注では、論理的な行動をとる人は決定可能な数学的命題について不確定でありえないと注意している。

ここで想定する個人は、通常の論理の基準を超えた基準が必要となるような状況で、意思決定を行う必要がある。したがって、ある規範が提示されたとき、

それに従って行動するか、あるいは、違う言い方をすれば、自分がそれに違反していることに気づいたらどう反応するかを自問自答しなければならない。経済理論では、家族、企業、国家などの組織に対しても個人に適用される理論が適用される可能性があるが、以下では個人という表現を用いる。

3.1.2 世界と世界の状態

§2.3 The world, and the states of the world では、意思決定の問題に関して以下の用語を用いるが、これらは統計学や日常の会話で用いられる用語と矛盾しない。

世界（world）	個人が関心を持つ対象
世界の状態（state）	世界に関する明確な記述
真の状態	実際の状態、すなわち世界に関する真の状態の記述

3.1.3 事象

§2.4 Events は状態の集合である事象についての記述である。集合論、確率論の用語と微妙な違いがあるため、念のために収録しておく。任意の与えられた

表 2.4.1. 状態と事象に関する表現

用語	定義
（基礎的な用語）	
集合	事象
$A, B, C, \dots$	事象を表す一般的な記号
s, s', s''	状態を表す一般的な記号
S	全事象
$\emptyset$	空事象
（関係）	
$s \in A$	状態 s は A の要素である
$A \subset B$ ($B \supset A$)	A の任意の要素は B の要素である
$A = B$	A は B と一致する
（構成）	
$\sim A$	S の要素のうち、 A に含まれない要素
$\bigcup_i A_i$ (和事象)	少なくとも一つの $A_1, A_2, \dots$ に含まれる S の要素
$A \cup B$	A と B の和事象（または合併）
$\bigcap_i A_i$ (共通事象)	すべての $A_1, A_2, \dots$ に含まれる S の要素
$A \cap B$	A と B の共通事象

世界に対して、全事象 (universal event) と空事象 (vacuous events) という、論理的に最も重要な二つの事象がある。全事象は、ここでは S と表して、世界のすべての要素を含む。「世界」が現実的、技術的な意味を持てば、 S は現実の正解となる。空事象は、ここでは 0 と表し、どの状態も要素に含まない。ある事象が生起したという表現で、その事象は真の状態を要素として含んでいることを表す。

3.1.4 結果、行為、意思決定

§2.5 Consequences, acts, and decisions では以下の概念が定義される。**結果** (consequences) とは個人に生じる可能性がある任意の出来事である。一般的に、結果は金銭、生活、健康状態、友人からの評価、他者の幸福、神の意志など、個人が関心を持つすべてを含む。世界の状態に対比させれば、結果は個人の状態と呼ぶこともできる。任意の問題で、想像できる結果の集合を F と表し、個々の結果を f, g, h , などと表す。

行為 (act) は、形式的には世界の状態 s に結果 $f(s)$ を対応させる関数であり、 $\mathbf{f}$ という記号を用いる。なお $\mathbf{f}$ という記号は、 $f(s)$ という通常の記号より、論理的に優れている。それは関数について「対数 (logarithm)」が「 x の対数 (logarithm of x)」というより優れているのと同じ理由であり、実際、後にそれらの区別が必要となる。いくつかの行為について同時に議論するときは、 $\mathbf{f}, \mathbf{g}, \mathbf{h}; \mathbf{f}, \mathbf{f}', \mathbf{f}''$; または $\mathbf{f}_1, \mathbf{f}_2$ のように異なる文字を使う。

3.1.5 行為の選好に関する単純順序

§2.6 The simple ordering of acts with respect to preferences では、行為の選好を考えるために、Savage は以下で展開する概念を導入する。

定義 要素 $x, y, z, \dots$ の集合に対して定義される関係 $\leq$ で、任意の x, y, z に対して次の性質を持つ関係を、**単純順序** (simple ordering) と呼ぶ。

1. $x \leq y$ または $y \leq x$ のいずれかがなりたつ。
2. $x \leq y$ かつ $y \leq z$ であれば、 $z \leq x$ である。

単純順序は、全順序から反対称律を外したものである。参考のために全順序の定義と比較すると 1. は完全律、2. は推移律であり、単純順序に含まれない 3 番

目の反対称律は $a \leq b$ かつ $b \leq a$ ならば $a = b$ を要請するものである。つまり、単純順序では、反対称性では排除される状態を許容する。この概念を用いて、「より選好されない」ことを表す記号として $\leq$ を単純順序とするのが次の公準 (postulate) である。

P1 関係 $\leq$ は行為に関する単純順序である。

注意すべきは P1 は世界の状態については明示的に参照していないことで、Savage [57] は、数学的な精緻化を除けば、世界の状態を参照しない限り追加的な公準を公式化することはできないと主張しているし、実際、これ以上の言及はない。P1 それ自体の意味する結論はそれほど豊富ではないが、次の定理は容易に導出される。

定理 2.6.1 行為の有限集合を $\mathbf{F}$ とするとき、 $\mathbf{F}$ に含まれるすべての $\mathbf{g}$ に対して、 $\mathbf{f} \leq \mathbf{g} \leq \mathbf{h}$ を満たすような $\mathbf{F}$ の要素 $\mathbf{f}$ と $\mathbf{h}$ が存在する。

定理 2.6.1 は、意思決定者である個人が有限な $\mathbf{F}$ において、他の任意の行為より選好されない行為 $\mathbf{h}$ が存在することを保証するという意味で、意思決定への応用において重要である。ときに無限個の行為を考えることが適当となるが、経済的な問題では、最低の金銭単位（1 ペニー）を分割できないことから、何らかの数学的な配慮が必要である。

関係 $\leq$ から導かれるいくつかの関係を表 2.6.1 に示す。単純順序の仮定 P1 は、それから導出される関係 $\geq, <, >$ 、および $\equiv$ の性質を定めている。

表 2.6.1. $\leq$ から導かれる関係	
導出される関係	定義
$\mathbf{f} \geq \mathbf{g}$	$\mathbf{g} \leq \mathbf{f}$
$\mathbf{f} < \mathbf{g}$	$\mathbf{g} \leq \mathbf{f}$ が成立しない
$\mathbf{f} > \mathbf{g}$	$\mathbf{g} < \mathbf{f}$
$\mathbf{f} \equiv \mathbf{g}$	$\mathbf{f} \leq \mathbf{g}$ かつ $\mathbf{g} \leq \mathbf{f}$
$\mathbf{g}$ は $\mathbf{f}$ と $\mathbf{h}$ の中間にある	$\mathbf{f} \leq \mathbf{g} \leq \mathbf{h}$ または $\mathbf{h} \leq \mathbf{g} \leq \mathbf{f}$

なお $\mathbf{f} < \mathbf{g}$ は「 $\mathbf{g}$ は $\mathbf{f}$ より選好される」、 $\mathbf{f} \equiv \mathbf{g}$ は「 $\mathbf{f}$ は $\mathbf{g}$ と同等 (equivalent) である」または「無差別 (indifferent) である」と表現する。

Savage は、P1 から導かれるいくつかの関係を練習問題としているが、重要な

ものも含まれているので、以下に、美添による証明とともに収録する。

Exercices ([57], p.19)

1. 関係 $\geq$ は単純順序である。

(証明) $\mathbf{f} \geq \mathbf{g}$ は $\mathbf{g} \leq \mathbf{f}$ と同等であり $\leq$ は単純順序だから、 $\geq$ も単純順序である。

2. 導出された関係 $\leq, \geq, <, >$ および $\doteq$ は**推移的**である。すなわち、単純順序の定義の 2 番目の性質を満たしている。

(証明) 推移率は単純順序の定義の一部だから、 $\leq$ と $\geq$ は推移的である。 $\doteq$ は $\leq$ かつ $\geq$ を意味するから推移的である。 $<$ については、 $\mathbf{f} < \mathbf{g}$ かつ $\mathbf{g} < \mathbf{h}$ であれば $\mathbf{f} \leq \mathbf{g} \leq \mathbf{h}$ となるが、ここでもし $\mathbf{f} \doteq \mathbf{h}$ であれば $\mathbf{f} \doteq \mathbf{g} \doteq \mathbf{h}$ となり、 $\mathbf{f} < \mathbf{g}$ に矛盾する。同様に、 $>$ も推移的である。

3. 行為の任意の対 $\mathbf{f}, \mathbf{g}$ に対して、三つの関係 $<, \doteq, >$ のいずれか一つだけが成立する。

(証明) 定義から $\mathbf{f} \leq \mathbf{g}$ または $\mathbf{f} \geq \mathbf{g}$ が成立する。ここで $\mathbf{f} < \mathbf{g}$ と $\mathbf{f} > \mathbf{g}$ が同時に成立することはない。もし $\mathbf{f} \doteq \mathbf{g}$ なら $\mathbf{f} < \mathbf{g}$ も $\mathbf{f} > \mathbf{g}$ も成立しない。したがって、三つの関係 $<, \doteq, >$ のいずれか一つだけが成立する。

4. もし $\mathbf{f} < \mathbf{g}$ かつ $\mathbf{g} \doteq \mathbf{h}$ であれば $\mathbf{f} < \mathbf{h}$ となる。

(証明) $\mathbf{f} \leq \mathbf{h}$ は明らかである。ここで $\mathbf{f} \doteq \mathbf{h}$ とすると $\mathbf{f} \doteq \mathbf{g} \doteq \mathbf{h}$ となって $\mathbf{f} < \mathbf{g}$ に矛盾する。

5. $\mathbf{f} \doteq \mathbf{g}$ であれば $\mathbf{g} \doteq \mathbf{f}$ である。

(証明) $\mathbf{f} \doteq \mathbf{g}$ も $\mathbf{g} \doteq \mathbf{f}$ も、二つの命題 $\mathbf{f} \leq \mathbf{g}$ および $\mathbf{g} \leq \mathbf{f}$ が同時に成立することと同等である。

6. 任意の $\mathbf{f}$ に対して $\mathbf{f} \doteq \mathbf{f}$ が成立する。

(証明) 任意の $\mathbf{f}$ に対して $\mathbf{f} \leq \mathbf{f}$ だから、 $\doteq$ の定義から $\mathbf{f} \doteq \mathbf{f}$ である。

7. 三つの行為 $\mathbf{f}, \mathbf{g}, \mathbf{h}$ のうち、少なくとも一つは他の二つの中間にある。どのような場合に中間の行為が二つ以上となるか。

(証明) $\mathbf{f}, \mathbf{g}, \mathbf{h}$ が構成する三つの対について、例えば $\mathbf{f} \leq \mathbf{g}$ か $\mathbf{g} \leq \mathbf{f}$ のいずれかが成立する。 $\mathbf{f} \leq \mathbf{g}$ かつ $\mathbf{g} \leq \mathbf{h}$ であれば $\mathbf{f} \leq \mathbf{g} \leq \mathbf{h}$ である。また $\mathbf{h} \leq \mathbf{g}$

であれば $\mathbf{g}$ と $\mathbf{h}$ を比較すればよい。後半の問いに関しては、一つの例は $\mathbf{f} \sim \mathbf{g} \leq \mathbf{h}$ の場合であり、このとき $\mathbf{f} \leq \mathbf{g} \leq \mathbf{h}$ と $\mathbf{g} \leq \mathbf{f} \leq \mathbf{h}$ の両方が成立する。

P1 については二つのまったく異なる解釈が可能である。一つは決定問題における人々の行動を予測するもの、もう一つは決定問題における論理に似た整合性の基準とみなすものであるが、ここでは第2の解釈だけが重要である。第1の解釈を経験的 (empirical)、第2の解釈を規範的 (normative) と呼ぶ。

行為の間の選好を半順序 (partial ordering) として分析する可能性、すなわち単純順序の定義の1.を、弱い命題「 $\mathbf{f} \leq \mathbf{f}$ 」で置き換えて、いくつかの行為の対は比較不可能とする誘惑はあるが、Savage は、それは内省的な優柔不断または無関心の表明であり、得るものは少ないとして否定的である。

3.1.6 確実性の原理

§2.7 The sure-thing principle において、ある事象 B が与えられたときの行為 $\mathbf{f}$, $\mathbf{g}$ の選好を議論するために、暫定的に**確実性の原理**という概念を導入する。それは、ある個人が B が生起したことを知っているときに $\mathbf{f}$ を $\mathbf{g}$ より選好し、また B が生起しなかったことを知っているときにも $\mathbf{f}$ を $\mathbf{g}$ より選好するなら、その個人は（無条件に） $\mathbf{f}$ を $\mathbf{g}$ より選好するというものである。

もう少し形式的には次のようになる。個人が事象 B が生起したことを知っているとき、または事象 $\sim B$ が生起したことを知っているときのいずれも $\mathbf{f}$ を $\mathbf{g}$ より選好しないなら、彼は $\mathbf{f}$ を $\mathbf{g}$ より選好しない。さらに（彼が B を不可能とみなさないという条件の下で）、彼が B が生起したことを知ったときには $\mathbf{g}$ を $\mathbf{f}$ より厳密に選好し、 B が生起しなかったことを知ったときには $\mathbf{f}$ を $\mathbf{g}$ より選好しないならば、彼は $\mathbf{g}$ を $\mathbf{f}$ より厳密に選好する。

確実性の原理を公準とすることについて、Savage は慎重である。それは知識および可能性という未定義の新しい概念を使っているため、これらの概念に関する公準がなければ数学的には役に立たないからであり、この点で公準 P1 と同様に受け入れることは不適切となる。したがって、この原理は、P1 をより明確にするような緩やかな公準とみなすことが望ましいと主張する。

そこで、「 B が生起したことを知ったときに $\mathbf{f}$ は $\mathbf{g}$ より選好される」という概念にどのような技術的な解釈を与えるべきかを考える。どのような合理的な解釈であっても、選好は B 以外の状態において $\mathbf{f}$ および $\mathbf{g}$ が取る値に依存してはならない。そこで、 $\mathbf{f}$ と $\mathbf{g}$ は B の外では互いに一致する、すなわち、すべての $s \in \sim B$ について $f(s) = g(s)$ と仮定しても一般性を失わない。この場合、 $\mathbf{f}$ と $\mathbf{g}$ は $\sim B$ が与えられたとき、すなわち B が生起しなかったことを知ったときに、同等とみなされる。そうすると、确实性の原理の最初の部分は次のように解釈される。 B の外では一致するように $\mathbf{f}$ と $\mathbf{g}$ が変更された後で、変更された $\mathbf{f}$ が変更された $\mathbf{g}$ より選好されないならば、 B が生起したことを知ったとき、 $\mathbf{f}$ は $\mathbf{g}$ より選好されない。この「 B の下で $\mathbf{f} \leq \mathbf{g}$ 」という概念を $\mathbf{f} \leq \mathbf{g} \text{ given } B$ と表記する。

$\mathbf{f} \leq \mathbf{g} \text{ given } B$ を定義するにあたって、二つの行為 $\mathbf{f}$ と $\mathbf{g}$ が $\sim B$ において一致するように変更された場合、選好の順序は実際に変更された内容に依存しないことが暗黙に想定されている。すなわち、二つの行為 $\mathbf{f}$ と $\mathbf{g}$ が $\sim B$ で一致し、かつ $\mathbf{f} \leq \mathbf{g}$ であれば、 $\sim B$ において一致するまま異なる値に変更した $\mathbf{f}'$ と $\mathbf{g}'$ についても $\mathbf{f}' \leq \mathbf{g}'$ が成立する。

同等な表現は次のようになる。 $\sim B$ において一致する二つの行為 $\mathbf{f}$ と $\mathbf{g}$ について、 $\mathbf{f} \leq \mathbf{g}$ とする。ここで $\mathbf{f}$ と $\mathbf{g}$ を $\sim B$ において変更した $\mathbf{f}'$ と $\mathbf{g}'$ も $\sim B$ において一致するならば、 $\mathbf{f}' \leq \mathbf{g}'$ が成立する。この仮定を形式的に表現するのが、次の公準 P2 である。

P2 行為 $\mathbf{f}, \mathbf{g}$ および $\mathbf{f}', \mathbf{g}'$ が次の条件

1. $\sim B$ において、 $\mathbf{f}$ と $\mathbf{g}$ が一致し、 $\mathbf{f}'$ と $\mathbf{g}'$ が一致する、
2. B において、 $\mathbf{f}$ と $\mathbf{f}'$ が一致し、 $\mathbf{g}$ と $\mathbf{g}'$ が一致する、
3. $\mathbf{f} \leq \mathbf{g}$,

を満たすとき、 $\mathbf{f}' \leq \mathbf{g}'$ となる。

「 $\leq \text{ given } B$ 」が単純順序となることは明らかである。また適当な修正を加えて、「 $\geq, <, >, \div \text{ given } B$ 」などの概念も定義される。次の性質も重要である。すべての $s \in B$ に対して $f(s) = g(s)$ となるとき、 $\mathbf{f} = \mathbf{g} \text{ given } B$ となる。

ここで、「時間に依存する」形で記述された決定問題を、時間に依存しない形で表現することは可能かつ有益である。個人は B が生じたかどうかを観測した後で、 $\mathbf{f}$ と $\mathbf{g}$ のどちらかを決断しなければならないとして、実際に B が生じたことを知ったときに、彼の決定はどうなるか。

時間と無関係には、個人はすべての $s \in B$ に対して $\mathbf{f}$ または $\mathbf{g}$ の結果を得ることができ、これとは独立にすべての $s \in \sim B$ に対して $\mathbf{f}$ または $\mathbf{g}$ の結果を得ることができる。ここで、 $s \in B$ に対しては、どちらの行為を選択するかが問題である。

時間と無関係かつきわめて形式的には、この個人は次の4つの行為について決定しなければならない。

$\mathbf{h}_{00}$ は B の上で $\mathbf{f}$ に一致し、 $\sim B$ の上で $\mathbf{f}$ に一致する、
 $\mathbf{h}_{01}$ は B の上で $\mathbf{f}$ に一致し、 $\sim B$ の上で $\mathbf{g}$ に一致する、
 $\mathbf{h}_{10}$ は B の上で $\mathbf{g}$ に一致し、 $\sim B$ の上で $\mathbf{f}$ に一致する、
 $\mathbf{h}_{11}$ は B の上で $\mathbf{g}$ に一致し、 $\sim B$ の上で $\mathbf{g}$ に一致する。

現在考えている問題は次の形となる。これら4つの行為のいずれも特定の $\mathbf{h}_{ij}$ より選好されないとすれば、 $i = 0$ または $i = 1$ か、すなわち、 $\mathbf{h}_{ij}$ は B の上で $\mathbf{f}$ と一致するか、あるいは B の上で $\mathbf{g}$ と一致するか。 $i = 1$ となり得る必要十分条件は $\mathbf{f} \leq \mathbf{g} \text{ given } B$ であることは容易に確認できる。実際、 $i = 1$ であれば $\mathbf{h}_{0j} \leq \mathbf{h}_{ij}$ となり、これは $\mathbf{f} \leq \mathbf{g} \text{ given } B$ を意味する。逆方向の議論では、もし $\mathbf{f} \leq \mathbf{g} \text{ given } B$ であれば、 $\mathbf{h}_{00} \leq \mathbf{h}_{10}$ かつ $\mathbf{h}_{01} \leq \mathbf{h}_{11}$ である。いま $\mathbf{h}_{10} \leq \mathbf{h}_{11}$ とすると、4つの可能性のいずれも $\mathbf{h}_{11}$ より選好されない。また $\mathbf{h}_{11} \leq \mathbf{h}_{10}$ とすると、すべての i, j に対して $\mathbf{h}_{ij} \leq \mathbf{h}_{10}$ である。以上が問題への解答である。

任意の $\mathbf{f}$ と $\mathbf{g}$ に対して $\mathbf{f} \leq \mathbf{g} \text{ given } B$ が成立する場合、個人の判断では B は実際上不可能 (virtually impossible) または B はゼロ事象 (null) である、という。(Savage は空事象を *vacuous event* と呼んで、*null event* と区別しているが、通常の確率論では空事象を *null event* と呼ぶので注意が必要である。以下では *vacuous event* を空事象、*null event* をゼロ事象と記す。)

ゼロ事象に関する簡単な性質は次の定理にまとめられる。最後の性質以外は

直観的な解釈が容易である。

定理 2.7.1

1. 空事象 0 はゼロ事象である。
2. B がゼロ事象であるための必要十分条件は、任意の f と g に対して $f \leq g$ given B が成立することである。
3. もし B がゼロ事象であり、かつ $B \supset C$ であれば、 C はゼロ事象である。
4. もし $\sim B$ がゼロ事象であれば、 $f \leq g$ given B となる必要十分条件は $f \leq g$ が成立することである。
5. $f \leq g$ given S となるための必要十分条件は、 $f \leq g$ が成立することである。
6. もし S がゼロ事象であれば、 f と g に対して $f \doteq g$ が成立する。

証明 (美添による)

1. すべての f と g に対して $f \leq g$ given 0 だから、 0 はゼロ事象である。
2. この命題はゼロ事象の定義と同等である。
3. すべての f と g に対して $f \leq g$ given B である。ここで $C \subset B$ だから、 $f \leq g$ given C が成立する。
4. $\sim B$ がゼロ事象であれば $f \leq g$ given $\sim B$ である。これと $f \leq g$ given B から必要条件として $f \leq g$ が導かれる。十分条件であることは明らか。
5. $\sim S = 0$ だから、 $f \leq g$ given S は $f \leq g$ と同等である。
6. もし S がゼロ事象であれば、すべての f と g に対して $f \leq g$ が成立する。したがって $f \doteq g$ となる。 ■

定理 2.7.1 の 6 番目の結論は病的な場合に対応しているので注釈が必要である。 S がゼロ事象で、すべての事象が現実的に不可能 (virtually impossible) という表現は直観的に理解しにくい、解釈としては個人は何が起こっても無関心ということである。 F の解釈を適当に制限すれば想像できるが、そのような状況には興味がないため、後に導入する公準 P5 によって可能性を排除する。

次の定理のために用語を定義する。なお定理 2.7.2 と系 2.7.1 の証明は美添による。

定義 B の分割 (partition) とは、有限個の事象 B_i の集合であり、 $i \neq j$ のとき $B_i \cap B_j = 0$ かつ $\bigcup_i B_i = B$ を満たすものである。

定理 2.7.2 B_i を B の分割として、それぞれの i に対して $\mathbf{f} \leq \mathbf{g}$ given B_i と仮定する。このとき $\mathbf{f} \leq \mathbf{g}$ given B が成立する。さらに、少なくとも 1 個の j について $\mathbf{f} < \mathbf{g}$ given B_j であれば、 $\mathbf{f} < \mathbf{g}$ given B が成立する。

証明 B が生起したとき、ちょうど一つの B_i が生起して $\mathbf{f} \leq \mathbf{g}$ given B_i となるから $\mathbf{f} \leq \mathbf{g}$ given B である。後半については B_j がゼロ事象ではないという条件が必要である。このときは、 $\mathbf{f} \geq \mathbf{g}$ は $\mathbf{f} < \mathbf{g}$ given B_j と矛盾するから $\mathbf{f} < \mathbf{g}$ given B が成立する。 ■

系 2.7.1 ゼロ事象の有限個の和はゼロ事象である。

証明 もし B の分割である B_i がすべてゼロ事象であれば、すべての $\mathbf{f}$ と $\mathbf{g}$ に対して、それぞれの i について $\mathbf{f} \leq \mathbf{g}$ given B_i となるから、定理 2.7.2 によって $\mathbf{f} \leq \mathbf{g}$ given B となる。したがって B はゼロ事象である。 ■

定数となるような行為、すなわちその結果 (consequences) が世界の状態に依存しない行為は重要である。特に、そのような行為の選好を利用して、結果に関する自然な選好が導かれる。

$\mathbf{f} \equiv \mathbf{g}$ は $\mathbf{f}$ が恒等的に g となること、すなわちすべての s に対して $f(s) = g$ となることを表す。ここで**結果の間の選好**を次のように定義することができる。任意の結果 g と g' について $g \leq g'$ とは、 $\mathbf{f} \equiv g$ かつ $\mathbf{f}' \equiv g'$ ならば $\mathbf{f} \leq \mathbf{f}'$ を意味する。

厳密な表現を省略しても、 $\mathbf{f} \leq g, g \leq \mathbf{f}$ given B などの意味は明らかだから、以下はこれらの表現を使用する。特に、 $f \leq g$ given B は自然な意味を持つが、それは後に導入する公準 P3 によって明確にされる。

この段階で、 $f(s)$ という記号を $\mathbf{f}$ の意味で用いることが不適切であることが明らかとなる。すなわち、 $f(s) \leq g(s)$ とは結果 $f(s)$ と $g(s)$ に関する記述であり、 $\mathbf{f} \leq \mathbf{g}$ は行為に関する記述である。そして以下では頻繁に両方の記述が必要となる。

ここで Savage は次の問題を考える。 $\mathbf{f} \equiv g$ かつ $\mathbf{f}' \equiv g'$ であり、 $g \leq g'$ と仮定しよう。このとき、ある B に対して $\mathbf{f} > \mathbf{f}'$ given B となることは認められるだろうか。その答えは専門用語をどのように解釈するかに依存するとして、後の記述でも言及される簡単な例を提示している。友人とピクニックに行こうとしている人が、テニスラケットまたは水着のどちらかを購入しようと考えているが、両方を買う余裕はない。もし、彼の決定である「ラケットの所有」と「水着の所有」を結果 (consequence) とみなすのであれば、彼の決定はピクニックが行われる場所に依存することになる。彼が水着を選好しているとしても、ピクニックの行き先が水辺から遠いことを知ったときは、その判断は逆になるだろうから、 $\mathbf{f} > \mathbf{f}'$ given B となりうる。しかし Savage が定式化しようとしている行為 (act) と結果 (consequence) に関しては、この解釈は正しくない。ラケットの所有と水着の所有は行為であって結果ではない。日常用語で言えば「所有に至る」または「購入する」という表現は行為を意味する。この問題における結果は、友人との楽しい水泳、友人が泳いでいる間炎天下で新品のラケットを持てあそぶこと、などである。このように最終的な結果まで考慮すれば、最初の問題に対する答えは否定的となる。次の公準 (P3) は事象に関する知識は結果に関する選好に影響を与えないこと、さらにゼロ事象でない限り、結果の間の選好関係は事象に関する知識とは無関係であることを表現している。

P3 いま $\mathbf{f} \equiv g$ かつ $\mathbf{f}' \equiv g'$ で B はゼロ事象ではないものとする。このとき、 $\mathbf{f} \leq \mathbf{f}'$ given B となる必要十分条件は $g \leq g'$ である。

次の定理はほとんど明らかである。

定理 2.7.3 B_i を B の分割とし、すべての i について $s \in B_i$ のとき $f(s) = f_i$, $g(s) = g_i$ ($f_i \leq g_i$) と仮定するとき $\mathbf{f} \leq \mathbf{g}$ given B が成立する。さらに、ある j に対して B_j がゼロ事象でないときは、 $\mathbf{f} < \mathbf{g}$ given B が成立する。

証明 (美添による)

すべての i に対して $\mathbf{f} \leq \mathbf{g}$ given B_i という仮定は、定理 2.7.2 によれば $\mathbf{f} \leq \mathbf{g}$ given B を意味する。後半部分も、同様にして定理 2.7.2 から導かれる。 ■

公準 P1 と P2 を前提とすれば、定理 2.7.3 は論理的に公準 P3 と同等であるが、Abraham Wald 以来の英米の統計学者（British-American School of statisticians）によって許容性（admissibility）の概念として広く受け入れられている。

3.2 個人確率

Savage [57] の 3 章は個人確率（Personal probability）と題されている。そこで Savage は、まず質的な個人確率、つづいて量的な個人確率を導入する。個人確率は、5 章の効用（Utility）とともに、主観的ベイズ統計学の中心的な話題である。定理 3.3.1 は容易であるが、定理 3.3.2 と定理 3.3.3 はかなり複雑なため、ここでは筆者による証明を追記している。

まず §3.1 Introduction において、いくつかの事象を例にして個人が判断する確からしさを表現させる方法に関する考察がある。個人の情報を抽出するさまざまな手段を比較したうえで、Savage が採用したのは意思決定の問題として個人の判断を調べる方法であり、例が提示されているが、より厳密かつ具体的な方法については、§3.2 Qualitative personal probability で解説する。

用語としては主観確率（subjective probability）、判断確率（judgmental probability）なども広く用いられるが、Savage は個人確率（personal probability）を採用しているので、ここでもそう表記する。

個人確率に関しては §3.2 で二つの公準 P4 および P5 が導入され、P1-3 とともに用いられる。そこではある事象が他の事象より確からしくないという概念に関する形式的な議論が導かれ、通常、確率の概念に対して数学的に与えられる定性的な性質が導出される。さらに数学的な確率の概念は §3.3 においてもう一つの公準 P6 を導入することによって数量化される。その後の節で、§3.4 は個人確率への批判に関連する数学的な議論、§3.5 は条件付個人確率、§3.6 は条件付に独立な観測値の長い系列による確実性への接近、§3.7 は個人確率の視点による独立な事象の列という概念の拡張を扱っている。

3.2.1 質的な個人確率

§3.2 Qualitative personal probability では質的（定性的）な個人確率を導入する。

まず、ある事象 A が生起したときに賞金 (prize) が得られる行為 $\mathbf{f}_A$ を次のように定義する。ただし $f' < f$ とする。

$$\begin{aligned} f_A(s) &= f & s \in A \text{ のとき} \\ f_A(s) &= f' & s \in \sim A \text{ のとき} \end{aligned} \quad (3.2.1)$$

次の公準は、賞金を前提として個人が選択する二つの事象について、その選択が賞金に依存しないことを表している。この公準は (3.2.1) 式に似た多くの定義を含むため、見かけ上は複雑である。

P4 事象を A, B として、結果 f, f', g, g' と行為 $\mathbf{f}_A, \mathbf{f}_B, \mathbf{g}_A, \mathbf{g}_B$ を以下のとおりとする。

1. $f' < f, \quad g' < g;$
- 2a. $f_A(s) = f, \quad g_A(s) = g \quad (s \in A \text{ のとき}),$
 $f_A(s) = f', \quad g_A(s) = g' \quad (s \in \sim A \text{ のとき}),$
- 2b. $f_B(s) = f, \quad g_B(s) = g \quad (s \in B \text{ のとき}),$
 $f_B(s) = f', \quad g_B(s) = g' \quad (s \in \sim B \text{ のとき});$
3. $\mathbf{f}_A \leq \mathbf{f}_B;$

このとき $\mathbf{g}_A \leq \mathbf{g}_B$ が成立する。

P4 を反映して、事象 A が B より確からしくないことを $A \leq B$ と記し、次のように定義する。

定義 $f' < f$ とするとき

$$\begin{aligned} f_A(s) &= f & s \in A \text{ (のとき)}, & \quad f_A(s) = f' & s \in \sim A \text{ (のとき)}, \\ f_B(s) &= f & s \in B \text{ (のとき)}, & \quad f_B(s) = f' & s \in \sim B \text{ (のとき)}; \end{aligned}$$

を満たす $\mathbf{f}_A, \mathbf{f}_B$ に対して $\mathbf{f}_A \leq \mathbf{f}_B$ となることが、 $A \leq B$ の必要十分条件である。

少なくとも一つの価値のある賞金が存在するという仮定は無害であり、そうでない状況は研究に値しないため、次の公準を提案する。

P5 少なくとも 1 対の結果 f, f' に対して $f' < f$ がなりたつ。

P1-5 から導かれる結果のすべては、以下で導入する定義と定理に反映される。

定義 事象同士の関係 $\leq$ で次の条件を満たすものを質的な確率 (qualitative probability) とよぶ。

すべての事象 B, C, D に対して, $\leq$ は次の条件を満たす。

1. $\leq$ は単純順序である。
2. $B \leq C$ となる必要十分条件は, $B \cap D = C \cap D = 0$ となる任意の D に対し $B \cup D \leq C \cup D$ となることである。
3. $0 \leq B$ および $0 < S$ となる。

定義の2番目の条件は, B と C のいずれも生起せず D が生起した場合に補償される賞金によって, B と C に関する個人の判断は影響されないことを主張している。

定理 3.2.1 事象に適用される関係 $\leq$ は質的な確率である。

証明 (美添による)

定義に記された条件を確認する。

1. P1 は $\leq$ が単純順序であることを主張しているから, 事象の関係 $\leq$ は単純順序である。すなわち, 任意の事象 A, B, C に対して, $A \leq B$ または $B \leq A$ のいずれかがなりたち, さらに $A \leq B$ と $B \leq C$ がなりたてば $A \leq C$ となる。
2. $f' < f$ として, 一般に, 事象 A の上で定義される関数 $f_A(s)$ を, $s \in A$ のとき $f_A(s) = f$, $s \in \sim A$ のとき $f_A(s) = f'$ とする。 $\mathbf{f}_B, \mathbf{f}_C, \mathbf{f}_{B \cup D}$ および $\mathbf{f}_{C \cup D}$ を同様に定義すると, $\sim D$ の上で $\mathbf{f}_B$ は $\mathbf{f}_{B \cup D}$ と一致し, $\sim D$ の上で $\mathbf{f}_C$ は $\mathbf{f}_{C \cup D}$ と一致する。また, D の上で $\mathbf{f}_B$ は $\mathbf{f}_C$ と一致し, D の上で $\mathbf{f}_{B \cup D}$ は $\mathbf{f}_{C \cup D}$ と一致する。 $B \leq C$ だから, $\mathbf{f}_B \leq \mathbf{f}_C$ である。ここで P2 から $\mathbf{f}_{B \cup D} \leq \mathbf{f}_{C \cup D}$ は $\mathbf{f}_B \leq \mathbf{f}_C$ と同等であり, $B \cup D \leq C \cup D$ となる。
3. S の分割 $S = B \cup \sim B$ に対して $\mathbf{f}_0 \leq \mathbf{f}_B$ given B かつ $\mathbf{f}_0 \leq \mathbf{f}_B$ given $\sim B$ である。定理 2.7.3 (p. 142) から $\mathbf{f}_0 \leq \mathbf{f}_B$ となり, したがって $0 \leq B$ となる。同様に定理 2.7.3 から $0 < S$ も導かれる。 ■

定理 3.2.1 からは, $\leq$ が「より確からしくない」を意味するのであれば, それから期待される通常の数値的確率に対して成立する数学的な帰結が導かれる。それらは以下の練習問題としているが, 形式的に証明されるだけでなく, 直観的にも解釈されるべきものである。

なお、練習問題としてはいないが、「 $B \doteq 0$ となる必要十分条件は B がゼロ事象となることである」という性質も導かれる。ただし、厳密にはこれは定理 3.2.1 だけからは導けない。

(美添によるこの性質の証明) $f' < f$ として、 $\mathbf{f}_B$ を、 $s \in B$ のとき $f_B(s) = f$ 、 $s \in \sim B$ のとき $f_B(s) = f'$ と定め、 $\mathbf{f}_0 \equiv f'$ と定める。 B がゼロ事象だとすると、定義からすべての $\mathbf{f}$ と $\mathbf{g}$ に対して $\mathbf{f} \leq \cdot \mathbf{g}$ given B が成立し、特に $\mathbf{f}_B \leq \cdot \mathbf{f}_0$ given B となるが、これは $B \leq \cdot 0$ を意味する。 $0 \leq \cdot B$ は明らかだから、 $B = 0$ が導かれる。

逆を示すためには、何らかの追加的な前提が必要である。ここでは結果のうちに最も選好される f^* と最も選好されない f_* の存在を仮定しよう。この仮定は Pratt-Raiffa-Schlaifer [54] などでも明示的に採用されているものである。 $f = f^*, f' = f_*$ においてあらためて $\mathbf{f}_0$ および $\mathbf{f}_B$ を定めると、すべての行為 $\mathbf{f}$ に対して $\mathbf{f}_0 \leq \cdot \mathbf{f} \leq \cdot \mathbf{f}_B$ given B がなりたつ。 $B \leq \cdot 0$ という仮定から $\mathbf{f}_B \leq \cdot \mathbf{f}_0$ given B だから、任意の $\mathbf{f}, \mathbf{g}$ に対して $\mathbf{f}_0 \leq \cdot \mathbf{f} \leq \cdot \mathbf{f}_B \leq \cdot \mathbf{f}_0 \leq \cdot \mathbf{g}$ given B 、すなわち $\mathbf{f} \leq \cdot \mathbf{g}$ given B となる。これは B がゼロ事象であることを表している。以上から、 $B \doteq 0$ は B がゼロ事象であることと同等である。

Exercises ([57], p.32, 以下, 3 と 5a のヒントを除く証明は美添による)

1. $B \subset C$ のとき、 $0 \leq B \leq C \leq S$ となる。

(証明) $B \subset C$ であればすべての s に対して $f_B(s) \leq f_C(s)$ となるから、 $\mathbf{f}_B \leq \mathbf{f}_C$ である。

- 2a. $B \cap D = C \cap D = 0$ とするとき、 $B < C$ の必要十分条件は $B \cup D < C \cup D$ である。

(証明) $B < C$ なら $B \geq C$ はなりたたないが、これは $B \cup D \geq C \cup D$ がなりたたないことと同等である。したがって $B < C$ は $B \cup D < C \cup D$ と同等である。

- 2b. $0 < C$ かつ $B \cap C = 0$ であれば、 $B < B \cup C$ となる。

(証明) 2a から、 $B = B \cup 0 < B \cup C$ である。

3. $B \leq C$ であれば $\sim C \leq \sim B$ である。逆もなりたつ。

(Savage による Hint): S を次の 4 つに分割する Venn 図を描け。 $B \cap C$,

$$\sim B \cap C, B \cap \sim C, \sim B \cap \sim C.$$

(証明) $B = (B \cap C) \cup (B \cap \sim C)$, $C = (B \cap C) \cup (\sim B \cap C)$, $\sim B = (\sim B \cap C) \cup (\sim B \cap \sim C)$,
 および $\sim C = (B \cap \sim C) \cup (\sim B \cap \sim C)$ である。これから $B \leq C \iff$
 $(B \cap C) \cup (B \cap \sim C) \leq (B \cap C) \cup (\sim B \cap C) \iff (B \cap \sim C) \leq (\sim B \cap C) \iff$
 $(B \cap \sim C) \cup (\sim B \cap \sim C) \leq (\sim B \cap C) \cup (\sim B \cap \sim C) \iff \sim C \leq \sim B$ となる。

4a. $B \leq C$ かつ $C \cap D = 0$ のとき, $B \cup D \leq C \cup D$ である。

(証明) $D \cap \sim B \subset D$ だから $D \cap \sim B \leq D$, したがって $B \cup D = B \cup (D \cap \sim B) \leq$
 $C \cup (D \cap \sim B) \leq C \cup D$ となる。

4b. $B \leq 0$ であれば, $B \cup C \doteq C$ かつ $B \doteq 0$ である。

(証明) $C \leq B \cup C \leq 0 \cup C = C$ から $B \cup C \doteq C$ となる。また常に $0 \leq B$ が
 なりたつから, $B \leq 0$ は $B \doteq 0$ を意味する。

4c. $S \leq B$ であれば, $B \cap C \doteq C$ かつ $B \doteq S$ である。

(証明) $C = S \cap C \leq B \cap C \leq C$ だから $B \cap C \doteq C$ となる。常に $B \leq S$ が
 なりたつから, $S \leq B$ は $B \doteq S$ を意味する。

4d. $B \cup D \leq C \cup D$ かつ $B \cap D = 0$ であれば, $B \leq C$ である。

(証明) $B \cup (D \cap \sim C) \leq B \cup D \leq C \cup D = C \cup (D \cap \sim C)$ であり,
 $B \cap (D \cap \sim C) = C \cap (D \cap \sim C) = 0$ だから, 質的な確率の定義の 2. より $B \leq C$ となる。

5a. もし $B_1 \leq C_1, B_2 \leq C_2$, かつ $C_1 \cap C_2 = 0$ であれば, $B_1 \cup B_2 \leq C_1 \cup C_2$ が
 なりたつ。

(Savage による Hint): B_2 および C_1 を, B_2, C'_1, Q を排反として, それぞ
 れ $B_2 = B'_2 \cup Q$ および $C_1 = C'_1 \cup Q$ と表し, 以下の計算の各段階を確認せ
 よ。 $B_1 \cup B'_2 \leq C_1 \cup B'_2 = C'_1 \cup B_2 \leq C'_1 \cup C_2$, したがって $B_1 \cup B_2 \leq C_1 \cup C_2$
 である。(美添による補足: 4a による)。

5b. $B_1 \cup B_2 \leq C_1 \cup C_2$ かつ $B_1 \cap B_2 = 0$ であれば, $B_1 \leq C_1$ または $B_2 \leq C_2$ が
 なりたつ。

(証明) もし $B_1 > C_1$ かつ $B_2 > C_2$ とすると $B_1 \cup B_2 > C_1 \cup C_2$ となり, 仮
 定と矛盾する。

6. $B \leq \sim B$ かつ $C \geq \sim C$ であれば, $B \leq C$ である。ここで結論における等号は仮定の両方で等号となる場合に限られる。

(証明) B', C', P, Q を互いに排反な事象として, $B = B' \cup P, C = C' \cup P, \sim B = C' \cup Q, \sim C = B' \cup Q$ とおく。ここで $P \leq Q$ または $Q \leq P$ のいずれかがなりたつ。もし $P \leq Q$ であれば, $B' \cup Q = \sim C \leq C = C' \cup P$ がなりたつためには $B' \leq C'$ でなければならない。一方, もし $Q \leq P$ であれば, $B' \cup P = B \leq \sim B = C' \cup Q$ がなりたつためには $B' \leq C'$ でなければならない。どちらの場合も $B' \leq C'$ がなりたつから, $B = B' \cup P \leq C' \cup P = C$ となる。

3.2.2 量的な個人確率

§3.3 Quantitative personal probability が [57] の中核をなす節である。前節の最後の練習問題では, 質的な確率が, 確率に与えられている通常の性質を持つことを示したが, 量的な確率を導入するためにはさらなる公準が必要である。一つの方向として de Finetti [42] が提案した S を任意に大きな数の同等な部分集合に分割する公準が考えられるが, Savage はそれには同意せず, その代わりに質的確率との関係を考慮して公準 P6 を提示する。公準 P6 を受け入れるなら, その後には質的確率に言及する必要はなくなる。なお, de Finetti の主張に関する詳細な議論はこの節には記されていない。

公準を提示する前に, しばらく難解な記述が続く。まず集合 S の上で定義される確率測度 (probability measure) としてそれぞれの $B \subset S$ に実数値を対応させる関数 $P(B)$ を, 次の条件を満たすものと定義するが, これは完全加法的ではなく, 有限加法的な確率測度である。

1. すべての B に対して $P(B) \geq 0$ となる。
2. $B \cap C = 0$ のとき $P(B \cup C) = P(B) + P(C)$ となる。
3. $P(S) = 1$ である。

S の上に確率測度 P と質的確率 $\leq$ が定義されているとき, 任意の B, C に対して, $P(B) \leq P(C)$ となる必要十分条件が $B \leq C$ であれば, P は $\leq$ と一致する, または厳密に一致する (strictly agree) と呼び, $B \leq C$ が $P(B) \leq P(C)$ を意

味するときは、 P は $\leq \cdot$ とほとんど一致する (almost agree) と呼ぶ。この用語は P が $\leq \cdot$ と（厳密に）一致するときは、 P は $\leq \cdot$ とほとんど一致するという意味で整合的である。さらに、 P が $\leq \cdot$ と一致すれば、 P の知識から $\leq \cdot$ の知識が導かれることも容易に示せるが、一方で、 P が $\leq \cdot$ とほとんど一致するだけであれば、後の §3.4 で例を示すように、 $B \leq C$ であっても $P(B) = P(C)$ となる場合があり、 P の知識からは $\leq \cdot$ に関する不完全な知識だけが導かれる。

定義 B の n 重のほとんど一様な分割 (n -fold almost uniform partition) とは、 B の n 個の要素への分割であって、任意の r 個の要素の和は、任意の $r+1$ 個の要素の和よりも確からしくないものをいう。この概念は他の箇所では用いられないが、簡単のため本稿では“ n -a.u.p.”と略記する。

定理 3.3.1 任意の大きな n に対して B の n -a.u.p. が存在するなら、すべての整数 m に対して B の m -a.u.p. が存在する。

証明 (Savage による) $n \geq m^2$ として、 $B_1, \dots, B_n$ を B の n -a.u.p. とする。ユークリッドの互除法を用いて n を $n = am + b$ と表す。ただし a と b は $m \leq a$ かつ $0 \leq b < m$ となる整数である。ここで $C_1, \dots, C_m$ を、それぞれの C_j が B_i の任意の a 個または $a+1$ 個の和となるような m 重分割とする。 $r < m$ とすると任意の r 個の C_j の和は、 ar 個から $(a+1)r$ 個の間の個数の B_j の和であり、 $r+1$ 個の C_j の和は $a(r+1)$ から $(a+1)(r+1)$ 個の間の個数の B_i の和である。ここで $r < m \leq a$ だから、 $(a+1)r = ar + r < ar + a = a(r+1)$ となる。したがって $C_1, \dots, C_m$ は B の m -a.u.p. である。 ■

定理 3.3.2 もし任意の大きな n に対して S の n -a.u.p. が存在するなら、 $\leq \cdot$ とほとんど一致する確率測度 P がただ 1 つ存在する。さらに、任意の ρ ($0 \leq \rho \leq 1$)、任意の $B \subset S$ と上で定義された一意的な P に対して、 $P(C) = \rho P(B)$ を満たす $C \subset B$ が存在する¹⁾。

証明 Savage の証明はいくつかの段階に分けて行われ、詳細はほとんど読者に

1) ここで Savage は次の点を注意している。加算加法的測度の場合の類推から、この結論は P が非原始的 (non-atomic) であることを意味すると予想されるかもしれないが、この予想は誤りである。

ゆだねられている。各段階は直前の結果が次の段階に利用されるように構成されている。5b を除いて、補足的な証明は美添による。実際、いくつかの段階の証明は容易ではなく、特に 5a と 8a の導出はかなり面倒である。Savage が想定していた証明は、もう少しきれいかもしれない。

1. 任意の正の n に対して、 S の n -a.u.p. が存在する。

(証明) 定理 3.3.1 で $S = B$ とおけばよい。

- 2a. $p_1, \dots, p_n$ は $0 \leq p_1 \leq p_2 \leq \dots \leq p_n$ かつ $\sum p_i = 1$ を満たす実数とする。
このとき次の式がなりたつ。

$$\sum_1^r p_i \leq r/n \quad (r = 1, \dots, n) \quad (1)$$

(証明) $n \sum_1^r p_i = r \sum_1^r p_i + (n-r) \sum_1^r p_i \leq r \sum_1^r p_i + (n-r)r p_r \leq r \{ \sum_1^r p_i + \sum_{r+1}^n p_i \} = r \sum_1^n p_i = r$

- 2b. さらに

$$\sum_1^{r+1} p_i \geq \sum_{n-r+1}^n p_i \quad (r = 1, \dots, n-1)$$

であれば、次の式がなりたつ。

$$\sum_1^r p_i \geq (r-1)/n \quad \text{および} \quad \sum_{n-r+1}^n p_i \leq (r+1)/n \quad (2)$$

(証明) $\sum_1^{r+1} p_i \geq \sum_{n-r+1}^n p_i = 1 - \sum_1^{n-r} p_i \geq 1 - (n-r)/n = r/n$ だから $\sum_1^r p_i \geq (r-1)/n$ となる。2 番目の不等式は、仮定 $\sum_{n-r+1}^n p_i \leq \sum_1^{r+1} p_i$ と、2a が $\sum_1^{r+1} p_i \leq (r+1)/n$ を意味することから導かれる。

- 2c. 任意の r 個の p_i の和は $(r-1)/n$ と $(r+1)/n$ の中間にある。

(証明) 任意の r 個の p_i の和は $\sum_1^r p_i$ と $\sum_{n-r+1}^n p_i$ の中間にある。ここで 2b を適用する。

- 2d. P が $\leq$ とほとんど一致すると仮定する。ここで $C(r, n)$ は、任意の n -a.u.p. のうちの r 個の和を表す記号とする。ただし、 n -a.u.p. は文脈によって必ずしも同じ分割ではない。このとき、次の式がなりたつ。

$$(r-1)/n \leq P(C(r, n)) \leq (r+1)/n \quad (3)$$

- (証明) まず, 定義から $C(r-1, n) \leq C(r, n) \leq C(r+1, n)$ がなりたつことに注意する。 B の n 個の要素への分割を $B_1 \leq B_2 \leq \dots \leq B_n$ として, p_i を $P(B_i) (i = 1, \dots, n)$ に割り当てると, $p_1 \leq p_2 \leq \dots \leq p_n$ である。さらに, $\sum_1^n p_i = \sum_1^n P(B_i) = P(\bigcup_1^n B_i) = P(S) = 1$ がなりたつ。また, $C(r, n) \leq C(r+1, n)$ だから, p_i は $P(C(r, n)) \leq P(C(r+1, n))$ を満たす必要がある。このような割り当ては可能であり, たとえば $p_i = 1/n$ はこの条件を満たす。 $\bigcup_1^{r+1} B_i$ は一つの $C(r+1, n)$ であり, $\bigcup_{n-r+1}^n B_i$ は一つの $C(r, n)$ だから, $\sum_1^{r+1} p_i = P(\bigcup_1^{r+1} B_i) \geq P(\bigcup_{n-r+1}^n B_i) = \sum_{n-r+1}^n p_i$ となる。したがって, p_i は 2a および 2b の仮定を満たす。任意の $C(r, n)$ に対して $P(\bigcup_1^r B_i) \leq P(C(r, n)) \leq P(\bigcup_{n-r+1}^n B_i)$ だから, 2c から (3) 式が導かれる。
3. $k(B, n)$ を $C(r, n)$ が B より確からしくならない最大の整数 r (ゼロになることもある) とする。関数 $k(B, n)$ は明確に定義され, $0 \leq k(B, n) \leq n$ を満たす。

(証明) 定義から $r = k(B, n)$ は一意的であり, $0 \leq r \leq n$ を満たす。

- 4a. P が $\leq$ とほとんど一致するならば, 次の式がなりたつ。

$$(k(B, n) - 1)/n \leq P(B) \leq (k(B, n) + 2)/n \quad (4)$$

(証明) $r = k(B, n)$ とすると, すべての $C(r+1, n)$ は $B < C(r+1, n)$ を満たす。さもないければ $r = k(B, n) \geq r+1$ という矛盾が生じる。ある $C(r, n)$ を C と書くと, $C \leq B$ となる。 $C \leq B < C(r+1, n)$ だから, $P(B) \geq P(C) \geq (r-1)/n$ かつ $P(B) \leq P(C(r+1, n)) \leq (r+2)/n$ である。

- 4b. 多くても一つの P だけが $\leq$ とほとんど一致する。

(証明) P と P' の両方が $\leq$ とほとんど一致すれば, 任意の n と $r = k(B, n)$ に対して, $(r-1)/n \leq P(B) \leq (r+2)/n$ かつ $(r-1)/n \leq P'(B) \leq (r+2)/n$ となるから, $|P'(B) - P(B)| \leq 3/n$ がなりたつ。このことはすべての B に対して $P'(B) = P(B)$ となることを意味する。

- 5a. B_i および C_i が $(n\text{-a.u.p. とは限らない})$ n 分割で, $B_1 \leq B_2 \leq \dots \leq B_n$ および $C_1 \geq C_2 \geq \dots \geq C_n$ のように並んでいるとき, 次の式がなりたつ。

$$\bigcup_{n-r}^n B_i \geq \bigcup_{n-r}^n C_i, \quad r = 0, \dots, n-1. \quad (5)$$

(証明) 大きな m に対して, S の mn -a.u.p. $D_1 \leq D_2 \leq \dots \leq D_N$ を考える。ここで $N = mn$ とする。まず, D_i の m 個の要素の和を D' とするとき, $D' \leq B_n$ となることを示す。もし $B_n < D'$ であればすべての B_i について $B_i < D' = C(m, N)$ が成立するから, それぞれの B_i について互いに排反な D' を考えれば, $S = \bigcup_1^n B_i < C(mn, N) = S$ となり, 矛盾する。同様にして, D_i の m 個の要素の和 D'' は $C_n \leq D''$ を満たす。これから $r = 0$ に対する (5) 式, $C_n \leq B_n$ が導かれる。

$r \geq 1$ については, まず, D_i のうちもっとも確からしい $m(r+1)$ 個の要素の和を $D^{(r)}$ とすると, $\bigcup_{n-r}^n B_i \geq D^{(r)}$ となることを示す。もし $\bigcup_{n-r}^n B_i < D^{(r)}$ であれば, ある j ($j \geq n-r$) に対して $B_j < C(m, N)$ となるから, $B_i < C(m, N)$ ($i = 1, \dots, n-r-1$) がなりたつ。その和が S となるような異なる $C(m, N)$ が存在するから, $S = \bigcup_1^n B_i < C(mn, N) = S$ となり, 矛盾する。同様にして, D_i のうちもっとも確からしくない $m(r+1)$ 個の要素の和を $D_{(r)}$ とすると, これは $C(m(r+1), N)$ であり, $\bigcup_{n-r}^n C_i \leq D_{(r)}$ となることが示される。以上から, $\bigcup_{n-r}^n C_i \leq D_{(r)} \leq D^{(r)} < \bigcup_{n-r}^n B_i$ がなりたつ。

5b. さらに, 二つの分割 $\{B_i\}$ と $\{C_i\}$ が, ほとんど一様な分割 (n -a.u.p.) であれば, 次の式がなりたつ。

$$\bigcup_1^r C_i \leq \bigcup_1^{r+2} B_i \quad (r = 1, \dots, n-2) \quad (6)$$

(Savage による証明) $\bigcup_1^{r+2} B_i \geq \bigcup_{n-r}^n B_i \geq \bigcup_{n-r}^n C_i \geq \bigcup_1^r C_i$.

5c. n -a.u.p. の r 個の要素の和は, 他の n -a.u.p. の $r+2$ 個の要素の和より確からしくない (not more probable)。

(証明) この結論は 5b から導かれる。

5d. $B \cap C = 0$ であれば, 次の式がなりたつ。

$$k(B, n) + k(C, n) - 2 \leq k(B \cup C, n) \leq k(B, n) + k(C, n) + 1 \quad (7)$$

(証明) $r_1 = k(B, n)$, $r_2 = k(C, n)$, $r = k(B \cup C, n)$ とする。もし $r_1 + r_2 \geq n - 1$ であれば右側の不等式は自明だから, $r_1 + r_2 \leq n - 2$ と仮定する。このとき, $B' = C(r_1 + 1, n)$ と $C' = C(r_2 + 1, n)$ で $B \prec B'$, $C \prec C'$ および $B' \cap C' = 0$ を満たすものが存在するから, $B \cup C \prec B' \cup C' = C(r_1 + r_2 + 2, n)$ となる。これから $r < r_1 + r_2 + 2$, 書き換えれば $r \leq r_1 + r_2 + 1$ と, 右側の不等式が得られる。

$B'' = C(r_1 - 1, n)$ と $C'' = C(r_2 - 1, n)$ で, $B'' \prec B$, $C'' \prec C$ かつ $B'' \cap C'' = 0$ となるものは常に存在する。このとき, $B'' \cup C'' = C(r_1 + r_2 - 2) \leq B \cup C$ であり, これから左側の不等式 $r_1 + r_2 - 2 \leq r$ が得られる。

6a. $C(r, m)$ が $C(s, n)$ より確からしくないとき, 次の式がなりたつ。

$$\left(\frac{r-2}{m} \right) \leq \left(\frac{s+2}{n} \right) + \frac{1}{mn} \quad (8)$$

(Savage による Comment): 一つの mn -a.u.p. を考えて, a.u.p. の任意の $t+2$ 個の要素の和は, 任意の t 個の要素の和より, 実際に, より確からしい (actually more probable) ことを利用せよ。

(証明) Savage の Comment については, もし $C(t, n) \geq C(t+2, n)$ であれば $C(t, n) \geq C(t+2, n) \geq C(t+1, n) \geq C(t, n)$ となり, $B_1 \doteq B_2 \doteq 0$ となるような二つの $C(1, n)$ が存在することになるが, そうすると $C(1, n) \geq 0 \doteq B_1 \cup B_2 = C(2, n)$ となる $C(1, n)$ が存在して, a.u.p. の定義に反する。

D_i ($i = 1, \dots, mn$) を mn -a.u.p. として, D_i の要素を n 個ずつまとめることにより m 個の集合を作ると, この集合は m -a.u.p. となることが示される。実際, D_i の rn 個の要素を含む $C(r, m)$ を考えると, 任意の $C(r+1, m)$ は $(r+1)n$ 個の要素を含むから $C(r, m) \leq C(r+1, m)$ となる。同様に m 個の D_i の要素を含む n -a.u.p. を構成する。このとき $C(s, n) \leq C(s+2, n) = C(m(s+2), mn)$ かつ $C(r, m) \geq C(r-2, n) = C(n(r-2), mn)$ だから, $C(r, m) \leq C(s, n)$ より $C(n(r-2), mn) \leq C(m(s+2), mn)$ となり, これが

ら $n(r-2) \leq m(s+2)+1$ となる。この不等式が成立することは、任意の a.u.p. について $C(r, n) \leq C(r', n)$ が $r \leq r' + 1$ と同等であることからわかる。最後に、両辺を mn で割れば、 $(r-2)/m \leq (s+2)/n + 1/mn$ が得られる。

6b. 次の不等式がなりたつ。

$$\left| \frac{k(B, m)}{m} - \frac{k(B, n)}{n} \right| \leq \frac{3}{m} + \frac{3}{n} + \frac{1}{mn} \quad (8)$$

(証明) $r = k(B, m)$, $s = k(B, n)$ とおくと、 $C(r-1, m) \leq B \leq C(r+1, m)$ および $C(s-1, n) \leq B \leq C(s+1, n)$ となるから、 $C(r-1, m) \leq C(s+1, n)$ および $C(s-1, n) \leq C(r+1, m)$ となる。これらに 6a を適用すると、 $(r-3)/m \leq (s+3)/n + 1/mn$ および $(s-3)/n \leq (r+3)/m + 1/mn$ となる。これらを書き換えると $r/m - s/n \leq 3/m + 3/n + 1/mn$ および $s/n - r/m \leq 3/m + 3/n + 1/mn$ となり、 $|s/n - r/m| \leq 3/m + 3/n + 1/mn$ が導かれる。

6c. 次の式で $P(B)$ を定義することができる。すなわち極限が存在する。

$$P(B) =_{\text{Df}} \lim_{n \rightarrow \infty} \frac{k(B, n)}{n} \quad (9)$$

(証明) 6b は、 $k(B, n)/n$ がコーシー列 (Cauchy sequence) であることを示しているから、その極限が存在する。

7. 上で定義された $P(B)$ は、 $\leq$ とほとんど一致する唯一の確率測度である。

(証明) まず、任意の B に対して $P(B) \geq 0$ は明らかである。性質の 5d は $k(B, n)/n + k(C, n)/n - 2/n \leq k(B \cup C, n)/n \leq k(B, n)/n + k(C, n)/n + 1/n$ を意味している。ここで $n \rightarrow \infty$ とすると、 $P(B) + P(C) \leq P(B \cup C) \leq P(B) + P(C)$ となり、 $P(B \cup C) = P(B) + P(C)$ が得られる。最後に、すべての n に対して $k(S, n)/n = 1$ だから、 $P(S) = 1$ がなりたつ。一意性は 4b によって保証される。

8a. 以下の性質を持つ、 B に含まれる集合の二つの無限列 C_n, D_n が存在する。

1. $C_n \cap D_n = 0$,
2. $C_n \subset C_{n+1}$ かつ $D_n \subset D_{n+1}$,
3. $P(C_n) \geq \rho P(B) - n^{-1}$,

$$4. P(D_n) \geq (1 - \rho)P(B) - n^{-1}.$$

（証明） $\rho = 0$ または $\rho = 1$ であれば結論は自明だから、 $0 < \rho < 1$ と仮定する。また、十分に大きな n について考えればよい。

$3n$ -a.u.p $B_1, B_2, \dots, B_{3n}$ を考える。また $r_1 = [3n\rho] - 1$ および $r_2 = [3n(1 - \rho)] - 1$ とする。ここで $[x]$ は x を超えない最大の整数を表す関数（floor function）であり、 $x - 1 < [x] \leq x$ を満たす。 $C_n = \bigcup_1^{r_1} B_i$ および $D_n = \bigcup_{3n-r_2+1}^{3n} B_i$ とおくと $C_n \cap D_n = \emptyset$ である。さらに、 (C_n, D_n, G_n) を B の分割とすると、 $G_n = C(r_3, 3n)$ は $r_3 = 3n - r_1 - r_2$ 個の要素を含み、 $2 \leq r_3 < 4$ である。 $P(B)(r_3 - 1)/3n \leq P(C(r_3, 3n)) < P(B)(r_3 + 1)/3n$ だから、 $P(B)/3n \leq P(G_n) < 5P(B)/3n$ となる。したがって $P(C_n) = P(C(r_1, 3n)) \geq P(B)(r_1 - 1)/3n = P(B)([3n\rho] - 2)/3n \geq P(B)(3n\rho - 3)/3n = P(B)(\rho - 1/n) > \rho P(B) - 1/n$ となる。同様に、 $P(D_n) = P(C(r_2, 3n)) \geq P(B)(r_2 - 1)/n \geq (1 - \rho)P(B) - 1/n$ もなりたつ。

C_{n+1} を構成するために、 G_n の $3n$ -a.u.p. を考えて、 C'_n をこの分割の $r_1 = [3n\rho] - 1$ 個の要素の和集合とする。上記の議論を C_n に適用すれば、 $P(C'_n) \geq P(G_n)(\rho - 1/n)$ が得られる。

$C_{n+1} = C_n \cup C'_n$ とおくと、 $P(C_{n+1}) = P(C_n) + P(C'_n) \geq \rho P(B) - 1/n + \{P(G_n)(\rho - 1/n)\} = \rho P(B) - 1/(n+1) + \{P(G_n)(\rho - 1/n) + 1/(n+1) - 1/n\} = \rho P(B) - 1/(n+1) + \{P(G_n)(\rho - 1/n) - 1/n(n+1)\}$ となる。ここで $P(G_n)(\rho - 1/n) \geq 1/n(n+1)$ となることを示す必要がある。 $P(G_n) \geq P(B)/3n$ かつ大きな n に対して $\rho - 1/n > 0$ だから、 $P(B)(\rho - 1/n)/3n \geq 1/n(n+1)$ を示せばよいが、これは $P(B)(\rho - 1/n) \geq 3/(n+1)$ と変形され、 n が大きいとき左辺は $P(B)\rho > 0$ 、右辺は 0 に近づく。同様に、 $(1 - \rho)P(D_{n+1}) \geq (1 - \rho)P(B) - 1/(n+1)$ も示すことができる。すべての大きな n に対する結果は、数学的帰納法を適用することによって得られる。

$$8b. P(\bigcup_n C_n) \geq \rho P(B), P(\bigcup_n D_n) \geq (1 - \rho)P(B), (\bigcup_n C_n) \cap (\bigcup_n D_n) = \emptyset \text{ がな}$$

りたつ。

(証明) $P(\bigcup_n C_n) = \lim_{n \rightarrow \infty} P(C_n) \geq \rho P(B)$; $P(\bigcup_n D_n) = \lim_{n \rightarrow \infty} P(D_n) \geq (1 - \rho)P(B)$. 明らかに $(\bigcup_1^n C_n) \cap (\bigcup_1^n D_n) = C_n \cap D_n = 0$ だから, $n \rightarrow \infty$ とすれば $(\bigcup_n C_n) \cap (\bigcup_n D_n) = 0$ が導かれる。

8c. $P(\bigcup_n C_n) = \rho P(B)$ である。

(証明) もし $P(\bigcup_n C_n) > \rho P(B)$ であれば, $P(B) \geq P((\bigcup_n C_n) \cup (\bigcup_n D_n)) = P(\bigcup_n C_n) + P(\bigcup_n D_n) > \rho P(B) + (1 - \rho)P(B) = P(B)$ となり, 矛盾する。したがって $P(\bigcup_n C_n) = \rho P(B)$ である。 ■

以上の定理では前提として「ほとんど一様な分割 (a.u.p.)」の存在を仮定している。その意味を考えるために, Savage はいくつかの概念を導入する。

定義 任意の $B \geq 0$ に対して, どの要素も B より確からしくない S の分割が存在する場合, $\leq \cdot$ を細密 (fine) と呼ぶ。

定義 B と C は, $B \cap G = C \cap H = 0$ を満たす任意の非ゼロ (non-null) 事象 G と H に対して $B \cup G \geq C$ かつ $C \cup H \geq B$ が成立する場合, ほとんど同等 (almost equivalent) と呼び, $B \approx C$ と表す。

同等な事象はほとんど同等であることは明らかである。

定義 ほとんど同等である任意の事象の対が同等となる場合, $\prec$ は緊密 (tight) である。

次の定理では「ほとんど一様な分割 (a.u.p.)」が, 細密 (fine) な $\leq \cdot$ の仮定から導かれる。この定理は段階的に示され, 3 と 5 については証明の概略が与えられているが, 詳しい説明はないため, 1, 2, 4, 6, 7 の全体, および 3, 5 への補足として, 美添の証明を記す。

定理 3.3.3 $\leq \cdot$ が細密であれば, 以下の性質がなりたつ。

1. $B > 0$ かつ $C > 0$ のとき, $0 \prec D \prec B$ となるような $D \subset C$ が存在する。
2. $B \approx G, C \approx H$ かつ $B \cap C = G \cap H = 0$ であれば, $B \cup C \approx G \cup H$ となる。
3. $B \approx C, G \approx H, B \cup C \approx G \cup H$, かつ $B \cap C = G \cap H = 0$ であれば, $B \approx G$ となる。

4. S のほとんど同等な事象 (almost equivalent events) への任意の分割は、ほとんど一様な分割 (a.u.p.) である。
5. すべての事象はほとんど同等な二つの事象に分割される。
6. 任意の非負の整数 n に対して、すべての事象はほとんど同等な 2^n 個の事象に分割される。
7. 次の命題が成立する。
 - (a) $\leq \cdot$ とほとんど一致する P が一意的に存在する。
 - (b) 任意の B, ρ ($0 \leq \rho \leq 1$) および (a) で定められた一意的な P に対して、 $P(C) = \rho P(B)$ となる $C \subset B$ が存在する。
 - (c) $B > \cdot 0$ のときは $P(B) > 0$ である。
 - (d) $B \approx \cdot C$ となる必要十分条件は $P(B) = P(C)$ である。

証明

1. S の分割でいずれの要素も B より確からしくないものが存在するから、その分割の要素のうちに $E \cap C > \cdot 0$ となる E が存在する。 $D = E \cap C$ とすると、 D は $D \subset C$ かつ $0 < \cdot D < \cdot B$ を満たす。
2. 非ゼロ事象 E を $(B \cup C) \cap E = 0$ を満たすものとして、 $E = E_1 \cup E_2$ と、二つの非ゼロ事象に分割する。このとき $B \cup E_1 \geq \cdot G$ かつ $C \cup E_2 \geq \cdot H$ だから、 $(B \cup C) \cup E = (B \cup E_1) \cup (C \cup E_2) \geq \cdot G \cup H$ となる。同様に、 $(G \cup H) \cap F = 0$ を満たす非ゼロ事象 F に対して、 $(G \cup H) \cup F \geq \cdot B \cup C$ である。以上から $B \cup C \approx \cdot G \cup H$ となる。
3. (Savage による証明、かっこ () 内が美添による補足である)
 結論を否定して、たとえば $B \cup E < \cdot G, B \cap E = 0, E > \cdot 0$ と仮定して、二つの場合を考える。
 (a) もし $B \cup C < \cdot S$ であれば、一般性を失うことなく、 $C \cap E = 0$ と仮定できるから $(B \cup C) \cup E \geq \cdot G \cup H$ であり、 $C > \cdot H$ となる。(なぜなら、もし $C \leq \cdot H$ であれば、 $(B \cup C) \cup E = (B \cup E) \cup C < \cdot G \cup H$ となって $B \cup C \approx \cdot G \cup H$ と矛盾する。)

E を二つの非ゼロ事象 E_1 と E_2 に分割すると, G の中に $C \cap E' = 0 < \cdot E' \leq \cdot E_2$ となるような E' がある。実際, もし C の外で G がゼロ (null) となる事象であれば $C \geq \cdot G > \cdot B \cup E$ となり, $C \approx \cdot B$ に反する。このとき $C \cup E' > \cdot H \cup E' \geq \cdot G > \cdot (B \cup E_1) \cup E_2$ だから, $C > \cdot B \cup E_1$ となり, $C \approx \cdot B$ に反する。(もし $C \leq \cdot B \cup E_1$ だとすると $C \cup E' \leq \cdot (B \cup E_1) \cup E' \leq \cdot (B \cup E_1) \cup E_2$ となって $C \cup E' > \cdot (B \cup E_1) \cup E_2$ に矛盾する。)

(b) もし $B \cup C \equiv S$ であれば, 簡単かつ特別な場合である $C \cap G \equiv 0$ を除いて, 以下の順に示される。 $H \cup G \equiv S$ (もし $H \cup G < \cdot S$ であれば空でない事象 E' が存在して $S > \cdot (H \cup G) \cup E' \geq \cdot B \cup C \equiv S$ となり矛盾する); $C < \cdot B \cup E < \cdot G$ ただし $E > \cdot 0$ かつ $E \subset C \cap G$ とする; $(B \cap H) \cup E < \cdot (G \cap C)$ ($B \cup E = (B \cap G) \cup (B \cap H) \cup E < \cdot G = (B \cap G) \cup (C \cap G)$ だから $(B \cap H) \cup E \leq \cdot C \cap G$ である); $(C \cap H) < \cdot (B \cap G)$ ($(C \cap H) \cup (C \cap G) = C < \cdot G = (B \cap G) \cup (C \cap G)$ だから $(C \cap H) < \cdot (G \cap B)$ である。さもなければ $C = (C \cap H) \cup (C \cap G) \geq \cdot (B \cap G) \cup (C \cap G) = G$ となって矛盾する); $H \cup E < \cdot G$ (p.147 の 5a によって, $H \cup E = (S \cap H) \cup E \equiv \{(B \cup C) \cap H\} \cup E = \{(B \cap H) \cup E\} \cup (C \cap H) < \cdot (C \cap G) \cup (B \cap G) = G$ となる)。以上で $B \cup E < \cdot G$ という仮定の矛盾が導かれた。同様にして $G \cup E < \cdot B$ もなりたたないから $B \approx \cdot G$ である。

4. $\{B_i\}$ が S の分割で $B_i \approx \cdot B_j$ であれば, $\{r \text{ 個の } B_i\} \approx \cdot \{r \text{ 個の } B_j\}$ となるから, $\{(r+1) \text{ 個の } B_i\} \geq \cdot \{r \text{ 個の } B_j\}$ である。
5. (Savage による証明, かっこ () 内が美添による補足である)

次の条件を満たす, B の分割 C_n, D_n, G_n が存在する。

1. $C_n \cup G_n \geq \cdot D_n, D_n \cup G_n \geq \cdot C_n,$
2. $C_{n+1} \supset C_n, D_{n+1} \supset D_n, G_{n+1} \subset G_n,$
3. $\sim G_{n+1} \cap G_n \geq \cdot G_{n+1}$; そのため G_n は少なくとも G_{n+1} と同等に確からしい二つの排反な事象を含む。

(B の分割 $A_1 \leq \dots \leq \cdot A_m$ を考えて, k を $\bigcup_1^k A_i \geq \cdot \bigcup_{k+1}^m A_i$ を満たす最小の整数とする。このとき $\bigcup_1^{k-1} A_i < \cdot \bigcup_k^m A_i$ となる。ここで $C_1 = \bigcup_1^{k-1} A_i, D_1 =$

$\bigcup_{k+1}^m A_i, G_1 = A_k$ とすると, これらは 1-3 の条件を満たしている。 C_n, D_n, G_n が条件 1-3 を満たすことを示すためには数学的帰納法を用いる。あらためて $A_1 \leq \cdots \leq A_m$ を G_n の分割とし, k を $C_n \cup (\bigcup_1^k A_i) \geq D_n \cup (\bigcup_{k+1}^m A_i)$ を満たす最小の整数とする。ここで $C_n \cup (\bigcup_1^{k-1} A_i) < D_n \cup (\bigcup_k^m A_i)$ に注意する。三つの事象 $C_{n+1} = C_n \cup (\bigcup_1^{k-1} A_i), D_{n+1} = D_n \cup (\bigcup_{k+1}^m A_i), G_{n+1} = A_k$ は条件 1-3 を満たす。)

任意の $H > 0$ に対して, いずれの要素も H より確からしくない m 個の事象への分割を考えて, n を $2^{n-1} > m$ となるように十分大きく定めると, $G_n < H$ となる。もし G_n が H よりも確からしいとすると, 分割のそれぞれの要素よりも確からしくなり, 分割のすべての要素の和, すなわち S は G_1 より確からしくなるから矛盾する。

($E_1 \leq \cdots \leq E_m$ を S の分割として, $G'_n = \sim G_n \cap G_{n-1}$ とすると, すべての i に対して $G_n > E_i$ となり, 任意の i, j に対して $G_{n-1} = G_n \cup G'_n > E_i \cup E_j$ となる。同様に $G_{n-2} = G_{n-1} \cup G'_{n-1} > \{ \text{任意の 4 個の } E_i \text{ の組合せ} \}$ となる。したがって, $G_1 > \{ \text{任意の } 2^{n-1} \text{ 個の } E_i \text{ の組合せ} \} = \bigcup_1^m E_i = S$)

二つの事象 $B_1 = \bigcup_n C_n$ と $B_2 = (\bigcup_n D_n) \cup (\bigcup_n G_n)$ が, 要求された B の分割となる。

6. 5 で定められた B の分割 B_1 と B_2 のそれぞれについてはほとんど同等な分割が存在し, 3 の結果から, 分割の 4 個の要素は互いにほとんど同等である。このような分割を n 回繰り返せば, 2^n 個のほとんど同等な要素が得られる。
7. (a) S を 2^n 個のほとんど同等な事象 $A_1, \dots, A_{2^n}$ に分割して, $P(A_i) = 1/2^n$ と定める。このような任意の分割の要素は A_i と同等であり, 2^{n+1} 個のほとんど同等な事象のうちの任意の二つの事象は, 一つの A_i と同等だから, この定義は一意的である。一般の B に関しては, r を $\{r \text{ 個の } A_i\} \leq B < \{(r+1) \text{ 個の } A_i\}$ として, $P(B) = \lim_{n \rightarrow \infty} r/2^n$ と定めると, この定義も一意的である。もし P と P' が $\leq \cdot$ とほとんど一致する確率測度であれば, 任意の n と B に対して $|P(B) - P'(B)| \leq 1/2^n$ となり,

これから $P(B) = P'(B)$ が得られる。

(b) $A_i^{(n)}$ ($i = 1, \dots, 2^n$) を B の 2^n 個のほとんど同等な事象として, C_n を $[\rho 2^n]$ 個の $A_i^{(n)}$ を含む事象とする。次に, $A_i^{(n)}$ を 2 個のほとんど同等な事象に分割して, 2^{n+1} 個のほとんど同等な事象 $A_i^{(n+1)}$ を得る。そうすると $C_{n+1} \supset C_n$ となり, $C = \bigcup_n C_n$ とすれば, $P(C) = P(\bigcup_n C_n) = (\lim_{n \rightarrow \infty} [\rho 2^n] / 2^n) P(B) = \rho P(B)$ となる。

(c) すべての i に対して $B_i \leq B$ となるような $S = \{B_i\}$ の分割を考える。 $P(B) = 0$ なら $P(B_i) \leq P(B) = 0$ となり, $P(S) = P(\bigcup_i B_i) = \sum_i P(B_i) = 0$ となるから, 矛盾である。

(d) (i) $B \approx C$ かつ $P(B) < P(C)$ と仮定すると, $P(B) < P(C) = \rho P(C)$ ($\rho < 1$) となる。 $\rho' < 1 - \rho$ となる ρ' に対して, $P(E) = \rho' P(C)$ となる事象 $E \subset C$ ($B \cap E = 0$) が存在するから, $P(B \cup E) = P(B) + P(E) = (\rho + \rho') P(C) < P(C)$ となり, $B \approx C$ という仮定, $B \cup E' \geq C$ すなわち $P(B \cup E) \geq P(C)$ に矛盾する。

(ii) $P(B) = P(C)$ なら, $P(B) \leq P(C)$ かつ $P(C) \leq P(B)$ だから $B \geq C$ かつ $C \geq B$ である。非ゼロ事象 E に対しては明らかに $B \cup E \geq C$ かつ $C \cup E \geq B$ だから, $B \approx C$ である。 ■

系 3.3.1 $\leq \cdot$ が細密 (fine) かつ緊密 (tight) であれば, $\leq \cdot$ とほとんど一致する確率測度は, $\leq \cdot$ と厳密に一致する。また, S を, 任意に大きな数の同等な事象に分割することができる。

証明 (前半) P は $\leq \cdot$ とほとんど一致するから, $B \leq C$ は $P(B) \leq P(C)$ を意味する。そこで, $P(B) \leq P(C)$ から $B \leq C$ が導かれることを示せばよい。

$P(B) = P(C)$ のときは, B と C は同等だから $B \leq C$ である。 $P(B) < P(C)$ のときは $B < C$ である (そうでなければ $B \geq C$ かつ $P(B) \geq P(C)$ となる)。いずれも $B \leq C$ である。

(後半) n を任意の整数とすると, $P(B_1) = (1/n)P(S) = 1/n$ を満たす B_1 が存在する。ここで $B'_1 = \sim B_1$ とすると, $1 = P(S) = P(B_1) + P(B'_1)$ よ

り $P(B'_1) = 1 - 1/n = (n-1)/n$ となる。 $B_2 \subset B'_1$ と $B'_2 = \sim(B_1 \cup B_2)$ を $P(B_2) = (1/(n-1))P(B'_1) = 1/n$ かつ $P(B'_2) = (n-2)/n$ となるように定める。この手順を $P(B_{n-2}) = 1/n$, $P(B'_{n-2}) = 2/n$ となるまで続けて、次に $P(B_{n-1}) = (1/2)P(B'_{n-2}) = 1/n$ となるように $B_{n-1} \subset B'_{n-2}$ を定めると、 $B'_{n-1} = \sim(B_1 \cup \dots \cup B_{n-1})$ に対して $P(B'_{n-1}) = 1 - (n-1)/n = 1/n$ となる。最後に $B_n = B'_{n-1}$ とする。この分割 $\{B_i\}$ は、求められた性質 $\bigcup_i^n B_i = S$ かつ $P(B_i) = 1/n$ を満たしている。 ■

定理 3.3.4 $\leq$ が細密かつ緊密となるための必要十分条件は、すべての $B \prec C$ に対してそれぞれの要素と B の和が C より確からしくない S の分割が存在することである。

証明 この仮定は、分割 $\{A_1, \dots, A_n\}$ が、すべての i に対して $A_i \cup B \leq C$ を満たすことを意味している。もし $\leq$ が細密かつ緊密なら、定理 3.3.3 によって、確率測度 P が存在する。 $\delta = P(C) - P(B) > 0$ とすると、系 3.3.1 から、 $P(A_i) < \delta$ を満たす分割 $\{A_i\}$ が存在する。このとき $P(A_i \cup B) < \delta + P(B) = P(C)$ であり、これから、すべての i に対して $A_i \cup B \leq C$ がなりたつ。

逆に、 $B \prec C$ に対しては、 $A_i \cup B \leq C$ となる分割 $\{A_1, \dots, A_n\}$ の要素はすべての i に対して $A_i \leq C$ を満たすから、 $\leq$ は緊密である。 $\leq$ が細密であることは次のように示される。任意の $B \approx C$ に対して、 $B \leq C$ または $B \geq C$ のうちの少なくとも一つがなりたつ。一方で、 $B \prec C$ も $C \prec B$ もなりたたない。なぜなら、 $B \prec C$ であれば、 $A_i \cup B \leq C$ となるような分割が存在し、 $B \approx C$ と矛盾する。同様に $C \prec B$ もなりたたない。したがって、 $B \leq C$ かつ $B \geq C$ であり、 B と C は同等である。定義によって、 $\leq$ は緊密である。 ■

Savage は、定理 3.3.3 および 3.3.4 の議論を考慮して、次の公準 P6' を暫定的に提示している。これは事象の間の関係 $\leq$ 、さらに行為の間の関係 $\leq$ を規定するものである。

P6' $B \prec C$ のとき、 S の分割で、各要素と B との和が C より確からしくない分割が存在する。

この公準に関して、Savage はかなり慎重に考えていることがうかがえて、 $\leq$ が細密かつ緊密とする公準 $P6'$ は、関連する文脈で de Finetti [23] や Koopman [40] が主張するように「 S が任意の個数の同等な事象に分割可能」と仮定するよりも正当化しやすいと指摘している。その根拠として、 $P6'$ は、論理的には分割可能性よりも強い結果を導くこととともに、以下のような解釈を示している。仮に、あなたが $B < C$ 、すなわち C からの利得が B からの利得より大きいと考えているとする。ここで、あなたの世界を、あなた自身が選ぶコインを n 回投げたときの各結果に対応させた 2^n 個の事象に分割する。Savage は、 B と特定の長さ n の列の和より、あいかわらず C を選ぶようにコインと十分大きな n を選択することは容易であると指摘する。ここで、コイン投げの表と裏の系列については同等に確からしいと仮定する必要はまったくない。さらに Savage は、Keynes [39] および Koopman [40] などの主張に言及し、特に彼らが $P6'$ が数値的な確率と質的確率の一致を意味することには同意しないだろうと記しているが、その内容の詳細は記されていない。

確率論に限定すれば公準 $P6'$ で十分であるが、後の第 5 章で効用を導入する際、確率を定義する行為の他にさらに一般的な行為を対象とすることを考慮に入れて、Savage は以下の、より強い公準 $P6$ を提示する。 $P6$ は明らかに $P6'$ を意味するが、 $P6'$ が受け入れやすいと考えるのと同じ理由で受け入れやすいと、Savage は主張している。なお $P6$ の原文を部分的に意識しているため、参考として原文を追記する。

P6 $g < h$ として f を任意の結果とすると、次の条件を満たす S の分割が存在する。すなわち、 g または h の片方を、分割の任意の一つの事象でのみ各要素 s に対して f と変更し、他の値は変更しないならば、変更された g と h の選好関係は変わらない。

P6 If $g < h$, and f is any consequence; then there exists a partition of S such that, if g or h is so modified on any one element of the partition as to take the value f at every s there, other values being unchanged; then the modified g remains less than h , or g remains less than modified h , as the case may require.

3.2.3 数学的な補足

§3.4 Some mathematical details では大きく二つの話題に触れている。一つの話題は、細密かつ緊密である、細密であるが緊密でない、細密でないが緊密である、細密でも緊密でもない、という4つの条件を満たすような質的確率が存在するかという問題を扱う。後に、いずれの解答も肯定的であることを示す。

もう一つは可測性の問題である。初等的な議論を除いて、通常的確率論では確率の対象を、すべての集合（事象）ではなく、可測な集合に制限する。また、排反事象の無限列が可測であれば、その和の確率は個々の確率の和となるという加算加法性（完全加法性）を仮定する。しかし、ここで展開する理論では、すべての集合に確率を定義せず、有限加法性のみを想定する。一般的な解答として、まず確率の対象を限定することには、初等的な説明を除いて技術的な困難はなくむしろ望ましいこと、次に加算加法性を公準とするより、いくつかの状況において特別な仮説とする方がよいとしている。

数学的な議論を紹介する前に、まず、測度に厳密に一致しないような質的確率が存在することを指摘する。実際、細密であるが緊密でない場合がその例である。一方で、有限な S の上の質的確率が常に厳密に一致する測度を持つか、という問題は未解決である。ただし、ほとんど一致する測度を持たない例は Kraft-Pratt-Seidenbert [41] が示している。

以下が数学的な議論である。まず単位区間のすべての部分集合に対して定義され、Lebesgue 測度が定義される集合に対しては Lebesgue 測度に一致し、任意の合同な集合に対して同じ測度を与えるような可算加法的な測度が存在しない例として、Savage は「Halmos [36] の Problem (b), p. 276 に記述がある」としているが、引用が違っているようで未確認である。一方で、Lebesgue 測度が定義される集合の上では Lebesgue 測度に一致し、任意の合同な集合に対して同じ測度を与えるような、有限加法的な測度は存在する。これは Banach [17] にある例（Dover 版 p. 20, フランス語版 p. 32）で、単位円周上の測度であり、Banach-Tarski の定理を思い起こさせる²⁾。したがって有限加法的な測度は可算加法的である

2) Dover 版の英訳は、Banach 空間の導入 (p. 33) で complete の定義が不完全であるなど、

必要はない。あらためて Lebesgue 測度の有限加法的な測度への展開を P として、 $B \leq C$ を $P(B) \leq P(C)$ で定義すれば、この質的確率は細密かつ緊密である。

緊密であるが細密でない質的確率の例のために、二つの排反な単位区間 S_1 と S_2 からなる S と、 S_1 と S_2 のそれぞれに有限加法的な Lebesgue 測度 P_1 と P_2 を導入する。この例の集合 B は $B_1 = B \cap S_1$ と $B_2 = B \cap S_2$ に分割される。そこで $B \leq C$ を $P_1(B_1) < P_1(C_1)$ 、または $P_1(B_1) = P_1(C_1)$ かつ $P_2(B_2) \leq P_2(C_2)$ と定める。このとき S をどちらも S_2 より確からしくない二つの事象に分割することは不可能だから、この $\leq$ は細密ではない。一方、 $\leq$ が緊密であることは容易に確認できる。

次に、上の例と同様、 S を測度 P_1 と P_2 が与えられた S_1 と S_2 の和として、 $B \leq C$ を $P_1(B_1) + P_2(B_2) < P_1(C_1) + P_2(C_2)$ 、または $P_1(B_1) + P_2(B_2) = P_1(C_1) + P_2(C_2)$ かつ $P_1(B_1) \leq P_1(C_1)$ の場合と定める。この $\leq$ は細密であるが緊密ではない。実際、 $P_1(B_1) + P_2(B_2) = P_1(C_1) + P_2(C_2)$ として $D > 0$ を $B \cap D = C \cap D = 0$ となる S_2 の部分集合とすれば、 $B \cup D > C$ かつ $C \cup D > B$ となり、 B と C はほとんど同等であるが、 $P_1(B_1) = P_1(C_1)$ でないかぎり、それらは同等ではない。

細密でも緊密でもない $\leq$ の例は容易であるとして、Savage は詳細を省略しているが、次のようにすれば構築できる。 S を有限加法的な Lebesgue 測度の拡張 P_1, P_2, P_3 が与えられた排反な単位区間 S_1, S_2, S_3 の和として、次のいずれか一つが当てはまる場合に $B \leq C$ と定義する。 $P_1(B_1) < P_1(C_1)$ のとき、または $P_1(B_1) = P_1(C_1)$ かつ $P_2(B_2) + P_3(B_3) < P_2(C_2) + P_3(C_3)$ のとき、または $P_1(B_1) = P_1(C_1)$ かつ $P_2(B_2) + P_3(B_3) = P_2(C_2) + P_3(C_3)$ かつ $P_3(B_3) \leq P_3(C_3)$ のとき。上の議論と同様にして、この $B \leq C$ は細密でも緊密でもないことが確かめられる。

すべての集合を可測とするべきかどうかという問題に関する感触を得るために、 S を単位立方体として、 $\leq$ と厳密に一致する測度として平行 6 面体の体積を P とする。すると平行 6 面体の有限個の和の確率は Jordan 測度となる。(こ

冗談のような誤りがあって、不安である。

ここで Savage は “Jordan content” という表現を用いているが、有限加法族における有限加法的な測度の意味に理解しておく。）平行 6 面体の有限個の和として表現できない集合に対する測度の定義は無限にある。理想化された個人は Borel 集合や Lebesgue 可測集合に対する確率を付与できると考えることは不自然ではないが、さらに進んで Lebesgue 非可測集合または Borel 非可測集合に関して真剣に比較しようとすることは現実的な意味をほとんど持たないと Savage は指摘する。

このあたりの説明は難解であるが、Savage は確率の対象とする事象には何らかの制限を導入する一方で、測度としては有限加法性のみを仮定するという選択をしている。すべての集合に対して確率を割り当てることが不都合な例として、Savage は「球は、それ自身と同じ球二つと分割合同である。」という Banach-Tarski パラドックスをあげているが、そこでは Zermelo の選択公理を用いて一定の体積をもたない集合が扱われる。

すべての集合（事象）の対を比較することを断念するのであれば、同じようにして、すべての行為の対を比較することも不適当な主張となる。Savage が [57] で扱う事象は S のすべての部分集合ではなく、Borel 集合族のような S 上の σ -集合代数（ σ 加法族、完全加法族）となる。すべての結果の集合は可測空間となり、ある特定の σ -加法族が選ばれる。そしてある行為は事象の可測空間から結果の可測空間への可測関数となる。

有限加法的な測度が中心的な考察の対象であるときに、有限加法族ではなく σ -加法族を要求するのは奇妙に思われるかもしれないが、可算な和の概念が必要であることは定理 3.3.2 の 8a–8c や、定理 3.3.3 の最後の部分の例でわかる。

現代の確率論では完全加法的な測度に依存することから、公準 P1–P6 に追加する公準として完全加法性を仮定する強い動機はあるとしながら、結局、Savage は、強い非整合性や非合理性がないかぎり公準とすべきでないとし、実際に、可算加法性の必要性を要求するような議論は知らないときまで主張する。多くの人（many of us）は、自然数、直線、平面の上の必然的に有限加法的な一様な確率密度を、自然な確率の問題とみなす、強い直観的な傾向を持っている。した

がって、可算加法性を公準として仮定するよりも、適用可能な場合に応じた特別な仮説として、広い分野における有用な定理を導く方がよい。以上が Savage の結論である。

中間的なまとめ

Savage [57] の 3 章は、本稿で紹介した内容の後、3 章では §3.5 質的および量的な条件付確率 (Conditional probability, qualitative and quantitative), §3.6 経験を通じた確実性への接近 (The approach to certainty through experience), §3.7 事象の対称な列 (Symmetric sequences of events) と続くが、この部分は次の稿で取り上げることにはしたい。

本稿では、Pratt [52] の後半部分と、Savage [57] の 1–3 章の途中までを解説し、あわせて Savage [59] における議論の一部を紹介した。いずれもベイズ統計学の本質にかかわる議論であり、現代的な統計学を考えるうえで見過ごすことのできない内容を含んでいる。

なお、[57] では、さらに 4 章 個人確率に対する批判的な論評 (Critical Comments on Personal Probability) と 5 章 効用 (Utility) まで公理的な議論が続くので、その部分も改めて解説したい。

謝 辞

本研究は JSPS 科研費 JP18H00837 の助成を受けた。

This work was supported by JSPS KAKENHI Grant Number JP18H00837.

以下の文献にはベイズ統計学の基礎として、今回は取り上げなかったものも収録している。

参考文献

- [1] 後藤智弘・後藤文廣・美添泰人「可変回帰係数モデルを用いた消費者行動の分析」『青山経済論集』第 63 巻第 3 号, pp. 45 – 74, 2011

- [2] 竹内 啓『歴史と統計学』日本経済新聞社, 2018
- [3] 美添泰人「ベイズの手法による統計分析 — 部分的なサーベイと今後の展望」, 東京大学出版会, 竹内 啓（編）『計量経済学の新展開』第 6 章, 1983
- [4] 美添泰人「ベイズの手法による 2 項回帰モデルの推定」東京大学出版会, 鈴木雪夫・竹内 啓（編）『社会科学の計量分析』第 3 章, 1987
- [5] 美添泰人「ベイズ多変量解析（その 1）」立正大学『経済学季報』第 38 巻, 1988
- [6] 美添泰人「ベイズ多変量解析（その 2）」立正大学『経済学季報』第 38 巻, 1988
- [7] 美添泰人「多重共線性へのベイズ・アプローチ」東京大学出版会, 鈴木雪夫・國友直人（編）『ベイズ統計学とその応用』第 7 章, 1989
- [8] 美添泰人「効用についてのパラドックス — ベイズ統計からの解答 —」立正大学『経済学季報』第 39 巻, 1990
- [9] 美添泰人「有限母集団からの標本抽出—ベイズ統計学からの解釈—」『創価経営論集』第 18 巻第 3 号, 1994
- [10] 美添泰人「ベイズの手法による分布ラグモデルと季節変動の分析」一橋大学『経済研究』第 45 巻第 2 号, 1994
- [11] 美添泰人「小地域統計の推定手法と応用」一橋大学『経済研究』第 52 巻第 3 号, pp. 231–238, 2001
- [12] 美添泰人「経済と統計の間で」『日本統計学会誌』第 39 巻シリーズ J 第 2 号, pp. 161–179, 2010
- [13] 美添泰人「統計的推論の原理とベイズの理論」『青山経済論集』第 63 巻第 3 号, pp. 113–143, 2011
- [14] 美添泰人「ベイズ統計学の原理と応用（その 1）」『青山経済論集』第 74 巻第 3 号, pp. 203–230, 2022
- [15] 渡辺澄夫『ベイズ統計の理論と方法』コロナ社, 2012
- [16] Akaike, H., “Likelihood and the Bayes procedure,” *Bayesian Statistics*, proceedings of the first international meetings held in Valencia (Spain). (Bernardo, J.M.*et al.*, eds), Valencia, University Press, 1980.
- [17] Banach, S., *Theory of Linear Operations*, Dover, 2009. (Savage の引用は *Théorie des opérations linéaires*, Funduszu Kultury Narodowej, 1932.)
- [18] Basu, D., “Recovery of ancillary information,” *Sankhya*, A. 26, 1964.
- [19] Bayes, Thomas, “An Essay towards solving a Problem in the Doctrine of Chances”. *Philosophical Transactions*. 53: 370–418, 1763. (<https://www.jstor.org/stable/105741> から入手可能)
- [20] Birnbaum, A., “On the foundations of statistical inference,” *JASA*, 1962.
- [21] Box, G. E. P. and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, 1973.
- [22] Cox, D.R., “The Choice Between Alternative Ancillary Statistics,” *JRSS*, B. 30,

- 1970.
- [23] De Finetti, B., “La prévision: ses lois, ses sources subjectives,” translated into English by Kyberg and Smokler [42].
 - [24] DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill, 1970.
 - [25] Dempster, A. P., “A Generalization of Bayesian Inference,” JRSS, B, 1968.
 - [26] Edwards, A. W. F., *Likelihood*, Cambridge University Press, 1972.
 - [27] Edwards, W., H. Lindman and L. J. Savage, “Bayesian statistical inference for psychological research,” Psychol. Rev. 70, 1963, reprinted in Savage [62].
 - [28] Fienberg, S. E. and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics*, North-Holland, 1975.
 - [29] Fisher, R. A., “Theory of statistical estimation,” *Proc. Camb. phil. Soc.*, 22, 700–725, 1925.
 - [30] Fisher, R. A., *Statistical Methods and Scientific Inference*, 3rd ed. Oliver and Lloyd, 1973.
 - [31] Fraser D. A. S., *Probability and Statistics, Theory and Applications*, Duxbury Press, 1976.
 - [32] Good, I. J., *Probability and the weighting of evidence*, Charles Griffin, 1950.
 - [33] Good, I. J., *The Estimation of Probabilities*, MIT Press, 1965.
 - [34] Hacking, I., *Logic of Statistical Inference*, Cambridge University Press, 1965.
 - [35] Hacking, I., *The Emergence of Probability*, Cambridge University Press, 1975.
 - [36] Halmos, P. R., *Measure Theory*, Springer, 1970, (Savage の引用は Van Nostrand, 1950.)
 - [37] Jeffreys, H., *Theory of Probability*, 3rd ed. (corrected), Clarendon Press, 1961, corrected impression 1967.
 - [38] Kalbfleisch, J. D. and D. A. Sprott, “Application of likelihood methods to models involving large numbers of parameters,” JRSS, B, 32, 1970.
 - [39] Keynes, J. M., *A Treatise on Probability*, Macmillan, 1921, second ed. 1929.
 - [40] Koopman, B.O., “The axioms and algebra of intuitive probability,” AMS, 1940.
 - [41] Kraft C. H., J. W. Pratt, and A. Seidenbert, “Intuitive probability on finite sets,” AMS, 1959.
 - [42] Kyberg, H. E. and H. E. Smokler, *Studies in subjective probability*, Robert Krieger, 1980, original material published by Wiley, 1964.
 - [43] Lehmann, E. L., “Significance level and power,” AMS, 29, 1958.
 - [44] Lehmann, E. L., *Testing Statistical Hypotheses*, Wiley, 1959.
 - [45] Lindley, D. V., “A statistical paradox,” Biometrika, 44, 1957
 - [46] Lindley, D. V., *Introduction to Probability and Statistics from a Bayesian View point*, Cambridge University Press, 1965
 - [47] Lindley, D. V., *Bayesian Statistics, a review*, SIAM Monograph, 1972.
 - [48] Lindley, D. V. and A. F. M. Smith, “Bayes estimates of the linear model (with

- discussion),” JRSS, B.3 4, 1972.
- [49] Mosteler, F. and J. W. Tukey, *Data Analysis and Regression*, Addison-Wesley, 1977.
 - [50] Pratt, J. W., “Length of confidence intervals,” JASA, 56, 1961.
 - [51] Pratt, J. W., “Shorter confidence intervals for the mean of a normal distribution with known variance,” AMS, 34, 1963.
 - [52] Pratt, J. W., “Bayesian interpretation of standard inference statements,” JRSS, B, 27, 1965.
 - [53] Pratt, J. W., H. Raiffa and R. Schlaifer, “The foundations of decision under uncertainty: an elementary exposition,” JASA, 63, 1964.
 - [54] Pratt, J. W., H. Raiffa and R. Schlaifer, *Introduction to Stalislical Decision Theory*, MIT Press, 1995.
 - [55] Raiffa, H. and R. Schlaifer, *Applied Stalislical Decision Theory*, MIT Press, 1961.
 - [56] Ramsey, F. P., “Truth and Probability,” 1931, reprinted in Kyberg and Smokier [42].
 - [57] Savage, L. J., *The Foundations of Statistics*, Wiley, 1954. Revised and enlarged edition, Dover Publications, Inc., 1972.
 - [58] Savage, L. J., “The foundations of statistics. Reconsidered,” Proc. Fourth Berkeley Sympos., 1961, reprinted in Savage [62].
 - [59] Savage, L. J. and other contributors, *The Foundations of Statistical Inference*, Methuen, 1962.
 - [60] Savage, L. J., “Bayesian Statistics,” in Recent Developments in Information and Decision Processes, (R. E. Manchol and P. Gray, eds.), Macmillan, 1962. reprinted in Savage [62].
 - [61] Savage, L. J., “Elicitation of Personal Probabilities and Expectations,” JASA, 1971, reprinted in Savage [62].
 - [62] Savage, L. J., *The Writings of Leonard Jimmie Savage — A Memorial Selection*, ASA/IMS, 1981.
 - [63] Tukey, J. W., “The future of data analysis,” AMS, 33, 1962.
 - [64] Tukey, J. W., *Exploratory Data Analysis*, Addison-Wesley, 1977.
 - [65] Wald, A., *Statistical Decision Functions*, Wiley, 1950.
 - [66] Yoshizoe, Y., “A Bayesian method of estimating mortality rates in small areas,” *Bulletin of the International Statistical Institute*, 2001.