第2章

導入：確信度，モデル，パラメータ

Contents

I just want someone who I can believe in,
（信じられる人を探しているの）
Someone at home who will not leave me grievin'.
（そばにいてくれる人を，私を悲しませない人を）
Show me a sign that you'll always be true,
（あなたがいつも正しいのなら証を見せて）
and I'll be your model of faith and virtue.
（あなたを信じられることを，あなたの素晴らしさを，私は示してみせるから）[1)]

本章の目的は，ベイジアン分析の概念的枠組みを紹介することである．ベイジアン分析は，2つの基本的な着想に基づいている．第一に，ベイズ推論とは複数の可能性に対して確信度を再分配することである．第二に，確信度を分配する対象となる可能性は，意味のある数学的モデルのパラメータである．これら2つの着想は，本書のあらゆる分析において概念的な基礎となっている．これらの着想のわかりやすい例を本章で紹介する．本書の他の部分は，単にこれら2つの着想の適用例を数学的・計算論的に述べているにすぎない．本章では，あらゆるベイジアン分析に共通している基本的な手続きについても説明する．

1)　本章では数学的モデル，パラメータ値の確信度，モデルの意味について紹介する．詩中では日常的な文脈で「モデル」，「信じる」，「正しい」という言葉を使っており，ベイズ的手法こそが（擬人化していえば）信じる相手であることをほのめかしている．（原文では，1行目は I just want someone who I can believe in と書かれているが，文法的には in whom I can believe が正しい．口語的な文体にするために，あえてこのように書いた．さらに文法的な正確性を追求するなら，原文のような強弱弱格ではなく，弱強格で書く必要がある．）

2.1. ベイズ推論とは確率の確信度を再分配すること

ある朝，家の前の歩道が濡れていたとしたら，その原因となるあらゆる可能性を考えるだろう．例えば，雨が降ったのかもしれないし，庭の散水があったのかもしれない．新たに地下水が噴出したのかもしれないし，下水管が破裂したのかもしれない．通行人が飲み物をこぼしたのかもしれないし，他にも可能性はあるだろう．もしこの時点で，歩道の一部が濡れていること以外に情報がなければ，既存の知識に基づいて，これらの可能性はいずれもある程度の事前の確信度を持つ．例えば，通行人が飲み物をこぼした可能性よりも，昨晩雨が降った可能性のほうが高いだろう．私たちは日々の生活の中で，新たな観測データを獲得し続けている．もし歩道だけでなく，街路樹や駐車車両までも見渡す限り濡れていたとしたら，雨が降ったという仮説に確信度を再分配する．通行人が飲み物をこぼしたというような他の可能性は，新たな観測事実を説明しないだろう．一方，もし濡れている範囲がごく限られており，数フィート先に空のカップが落ちていたとしたら，たとえ雨が降ったことにくらべて可能性が低いとしても，通行人が飲み物をこぼした可能性に確信度を再分配するだろう．このように，様々な可能性に確信度を再分配することが，ベイズ推論の本質である．

架空の探偵シャーロック・ホームズの不朽の名言もまた，ベイズ推論の例としてふさわしい．ホームズはたびたび相棒のワトソン医師に，「何度も言っているだろう？　ありえないことをすべて取り除いて，それでもなお何かが残るのであれば，それが真実に違いないさ．たとえ，どんなに可能性が低そうに思えてもね」(Doyle, 1890, 6 章) と述べる．ホームズもワトソンも作者のドイルですらも，この推理がベイズ推論であるとは言及していない．しかし，これはベイズ推論なのである．ホームズは犯罪が起こると，複数の原因を考える．その中には，あらかじめ不可能と考えられる可能性も含まれている．ホームズは，多くの可能性を排除できるような証拠を体系的に集める．もしたった 1 つの可能性を除き，その他の可能性がすべて排除されたとしたら，残った可能性がありえないように思えたとしても，（ベイズによる）推理により完全に確信を持たざるをえない．

図 2.1 はホームズの推理を表している．図で表しやすいように，ある結果を説明する原因が 4 つに絞られている状況を想定した．それぞれ A，B，C，D とラベルがつけられている．グラフ内における棒の高さは，それぞれの可能性の確信度を示している（「確信度」は「確率」と同義である．ここでは日常的な用語として「確信度」を用いているが，後に数式で記述する際に「確率」という用語も用いる）．確信度は 0 から 1 の範囲をとる．もしある原因の候補の確信度が 0 であったなら，その原因は結果の説明としては不適当である．もし，ある原因の候補の確信度が 1 であったなら，その原因こそが結果の説明として適当である．原因の候補どうしは排他的であり，あらゆる原因が考慮されていると仮定するため，原因の確信度の総計は 1 となる．

図 2.1 の左上は，4 つの原因の候補はどれも事前の確信度が同等，つまり 0.25 であるこ

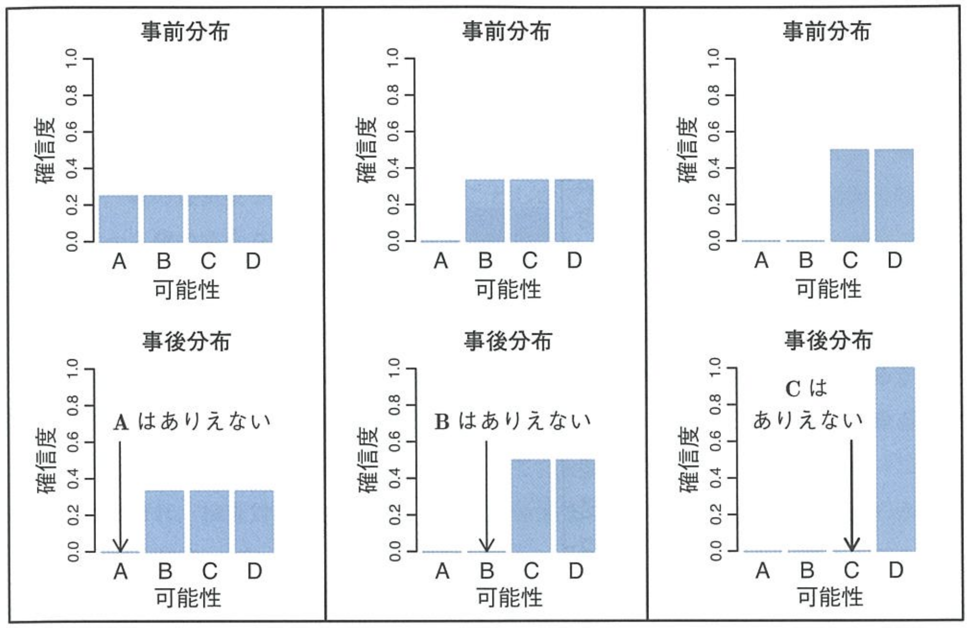

図 2.1　左上のグラフは，ある結果に関する 4 通りの原因の確信度を示している．原因にはそれぞれ A，B，C，D とラベルがつけられている．これらは排他的であり，あらゆる原因が考慮されている．初めは，すべての原因は等しく確信できるので，事前の確信度はどれも 0.25 である．左下のグラフは，原因の 1 つがありえないことが判明した場合の確信度を示している．結果として得られた事後分布は，中央の列における事前分布として用いられている．ここでは別の原因もありえないことが判明した．さらに中央の列における事後分布は，右の列の事前分布として用いられている．確信度を再分配するベイズの方法により，残りの可能性こそが原因であると示唆される．

とを示している．濡れた歩道の例では，雨が降ったという可能性は，新たに地下水が噴出した可能性よりも高いという事前の知識があった．しかしここで図示した例では，原因の候補に関する事前の確信度はどれも同等であると想定している．ここで，原因の候補 A を排除する新たな観測データを得たとしよう．例えば，仮に A が容疑者であったとき，A が犯行当時に現場から離れた場所にいたという情報を得たとする．すると，図 2.1 の左下のように，原因の候補 B，C，D に対して確信度を再分配しなければならない．確信度が再分配された分布は，新たな観測データを考慮した後の信念であるため，事後分布と呼ばれる．事後分布では原因 A の確信度は 0 となり，残った原因の候補 B，C，D に対して確信度が 0.33 ずつ（すなわち 1/3 ずつ）再分配されている．

事後分布は，さらなる観測の事前の信念となる．よって，図 2.1 中央列上段の事前分布は，左列下段の事後分布である．ここで，新たに獲得した証拠が原因の候補 B を排除したとしよう．すると図 2.1 中央列下段のように，残った原因の候補 C，D に確信度が再分配される．この事後分布は，図 2.1 の右上に示したように，さらなるデータ獲得における事前分布となる．最後に，もし新たなデータが原因の候補 C を排除したとしたら，図 2.1 の右下に示したように，すべての確信度は残った原因 D へ集約される．まさにホームズの宣言通りである．確信度の再配分は直観的なだけでなく，後に述べるように，ベイズ推論が数学的に行っていることとも合致するのである．

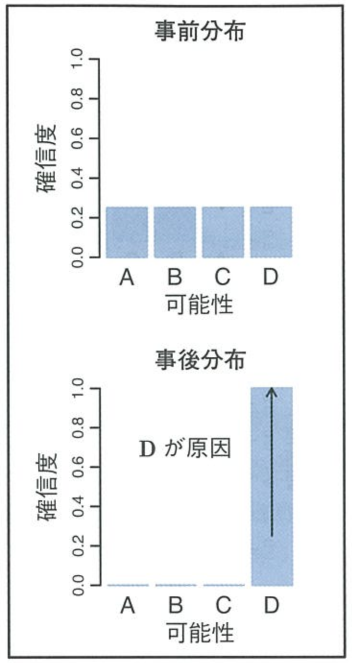

図 2.2 上のグラフは，ある結果に関する 4 通りの原因の確信度を示している．原因にはそれぞれ A，B，C，D とラベルがつけられている．これらは排他的であり，あらゆる原因が考慮されている．初めは，すべての原因は等しく確信できるので，事前の確信度はどれも 0.25 である．下のグラフは，どれが原因であるかが判明した場合の確信度を示している．確信度を再配分するベイズの方法により，その他の原因は「無罪放免」となっている（原因であることに対する確信度が 0 になっている）．

補完的な推理の方法もまたベイズ的であり，これは司法においては無罪放免と呼ばれる．ある犯罪に複数の容疑者が挙がっており，容疑者どうしは互いに無関係で，他に考えられる容疑者はいないとしよう．もしある容疑者が完全に疑わしいという証拠が見つかったとすると，他の容疑者は無罪放免となる．

この無罪放免の事例を図 2.2 に示した．上部は，ある結果に対して A，B，C，D の 4 つの原因が存在することを仮定している．これらの原因の可能性は排他的であり，他に考えられる原因は存在しないとする．犯罪の容疑者の文脈では，容疑者 A が犯人であるという仮説の確信度は，その容疑者が疑わしい程度を示している．よってこの文脈では，確信度よりも疑わしさについて考えるほうがわかりやすいだろう．4 人の容疑者の事前の疑わしさは同等であるため，図 2.2 の上部では 4 本の棒はすべて同じ高さ (0.25) になっている．ここで新たに，容疑者 D が犯人であることを示唆する有力な証拠があがったとする．他の容疑者は D と共犯関係にはないため，図 2.2 の下部の通り無罪放免となる．ホームズ流の推理のように，この無罪放免は直観的なだけでなく，後に述べるようにベイズ推論が数学的に行っていることとも合致する．

2.1.1. データにはノイズが含まれ，推論は確率的である

図 2.1 や図 2.2 の事例では，観測されたデータは，原因の候補に対して決定的な関係に

あることを想定していた．例えばシャーロック・ホームズが犯行現場で足跡を見つけ，その靴のサイズや種類に関して確信を得たとしよう．このときホームズは，ある容疑者が犯人であるか否かに関して結論を出すことができる．もちろん現実では，データとその背後にある原因は，確率的な関係にある．現実の刑事は足跡の大きさや靴底の模様を慎重に調べても，靴の選択肢を確率的に絞り込むことしかできないだろう．測定は完璧ではないし，足跡から靴の完全な情報が得られるわけではない．原因（すなわち靴）と測定された効果（すなわち足跡）の関係はランダムに変動する[a]．

科学的研究では，測定にランダムさはつきものである．どれだけ無関係な影響を排除しようとしても，執拗に測定に入り込んでくる．例えば，新薬がヒトの血圧を下げるかどうかを検証したいとしよう．何人かの検査対象者を投薬群に，別の何人かの検査対象者はプラセボ[b]を与える統制群に，ランダムに割り当てたとする．このような手続きは「二重盲検法」と呼ばれ，検査対象者も実験者も，誰が新薬またはプラセボを与えられたのかを知りえないようになっている（この情報はコンピュータによりランダムに割り当てられたコードに基づいており，コードの解読はデータ集計後に行われる）．数日間，毎日決まった回数だけ検査対象者の血圧を測り続けたとしよう．もちろん，運動やストレス，最近摂取した食べ物など，様々な要因によって血圧にはかなりのばらつきが生まれる．血圧の測定自体は，袖の上から血流の音を検出できるかどうかに依存しているので，不確実性を伴う．さらに，血圧は個人差が非常に大きい．その結果として，データは各群内でも群間でも，非常にばらつきが大きくなってしまう．よって，投薬群の中にはプラセボ群よりも血圧が高い検査対象者が存在しうるし，逆もまた然りである．このように，分散し重複する数値の山が 2 つあるとき，血圧の群間差がどの程度認められるか，その差はどの程度確からしいかを知りたい．問題は，実際に新薬に効果があったとしても，測定できる効果はランダムにしか見えないことである．

あらゆる科学的データには，多かれ少なかれ「ノイズ」が付随している．データ分析のテクニックは，ノイズが含まれるデータの背後に潜む傾向を推測するためにデザインされている．シャーロック・ホームズなら，観測により原因の候補のうちいくつかを完全に排除できる．しかし私たちはデータを集めたとしても，可能性の確信度を徐々に調整していくしかない．本書では，現実的な例をいくつも紹介する予定である．ベイジアン分析の素晴らしいところは，現実に問題となるような状況においても，確信度をどの程度再分配すべきかを数学的に解明できる点である．

データにノイズが含まれている場合におけるベイズ推論を，単純な例で説明しよう．ある工場では，直径が 1.0，2.0，3.0，4.0（デシメートル[c]のように，何らかの単位だとす

a) 訳注：random とは，「デタラメに」とか「バラバラな」などの意味があるが，統計的には「確率的に」という意味でもある．

b) 訳注：プラセボは偽薬ともいわれ，薬としての効果を何も持っていないが外見は薬のように見えるもの．本物の薬の効果はプラセボとの対比で検証される．

c) 訳注：デシメートル（dm）は 1/10 メートルを表す単位．

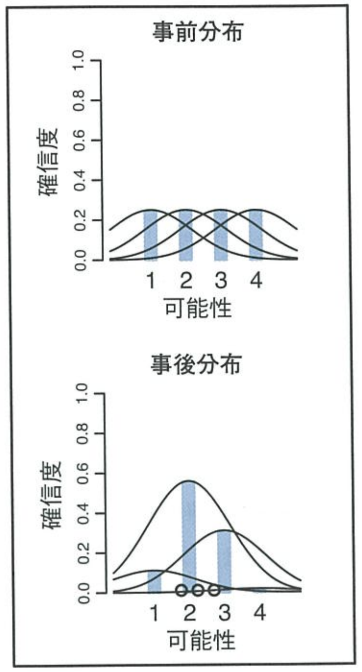

図 2.3　上のグラフは，正規分布における平均値に関する，事前の確信度を表している．平均値は 1，2，3，4 の 4 通り存在する．それぞれに対応する正規分布が重ね描かれている．横軸は，ボールのサイズ（青い棒が対応している）と，測定された直径（釣り鐘型の分布で表されている）の両方を示している．図の下では，実際に測定された 3 つのデータを横軸上に円で示している．4 通りの平均値の候補に確信度を再分配することで，データに基づくとサイズ 2 のボールがもっとも確信でき，サイズ 3 のボールは幾分かは確信できる，などのことがわかる．

る）のボールを製造している．ただし，同じサイズのボールであっても，ボールを膨らませる程度はランダムであり，製造過程にはかなりの違いがあるとする．よって，サイズ 3 として製造されたボールは，直径の全体的な平均が 3.0 であっても，個別のボールの直径は 1.8 から 4.2 までばらついているかもしれない．この工場にサイズ 2 のボールを 3 つ発注したら，直径がそれぞれ 1.77，2.23，2.70 のボールが納品されたとしよう．このような測定結果が得られたとき，この工場は正しくサイズ 2 のボールを納品したといえるだろうか？　あるいは，誤ってサイズ 1 や 3，サイズ 4 のボールまでも納品してしまったのだろうか？

この問いに対するベイズ的な答えを図 2.3 に示している．上部のグラフにおける青い棒を見ると，ボールのサイズが 1，2，3，4 のいずれの場合でも，事前の確信度は等しく 0.25 である．これは，工場は注文内容を忘れてしまい，どのサイズのボールが納品されてもおかしくないだろう，という想定に相当する．

この時点では，ボールの直径に関するランダムな変動が，どのように生じるかを特定しなければならない．図で説明しやすいように，ボールの直径の平均は 1.0，2.0，3.0 または 4.0 であるが，個別の膨張程度により，より大きくなることもより小さくなることもありうるとしよう．図 2.3 における釣り鐘型のカーブは，各ボールサイズの直径の確率を表している．すなわち，釣り鐘型のカーブがサイズ 2 に集中していたら，多くの場合は直

径 2.0 のボールが製造されるが，ランダムな膨張度合いにより，直径が 2.0 より大きくなることも小さくなることもある．図 2.3 の横軸は，ボールのサイズ（青い棒が対応している）と，測定された直径（釣り鐘型の分布で表されている）の両方を示している．

図 2.3 の下段は，実際に測定されたボールの直径を横軸上に円で示している．実際の直径はサイズ 2 やサイズ 3 に近いことがわかるだろう．しかし釣り鐘型の分布を見ると，たとえサイズ 1 のボールであってもこの程度の直径になりうることがわかる．すなわち，直観的にはもっとも確信できるボールのサイズは 2 であるが，データを見るとサイズ 3 の可能性もあり，サイズ 1 の可能性もわずかに捨てきれない．しかしサイズ 4 である可能性は低いと考えてよいだろう．図 2.3 の下段の通り，これらの直観はベイジアン分析に正確に反映されている．青い棒の高さは，4 通りのボールサイズに再分配された確信度を表している．データから，ボールはサイズ 2 である可能性が 56%，サイズ 3 である可能性が 31%，サイズ 1 である可能性が 11%，そしてサイズ 4 である可能性は 2% しかないことがわかる．

個々のボールの直径という「ノイズの含まれる」データから，本来製造されるはずのボールサイズを推論することは，現実の科学的研究や適用事例におけるデータ分析に酷似している．データは，背後にある発生メカニズムを知るための，ノイズを含んだ手がかりとなる．発生メカニズムに関していくつかの仮説を立てて，データによってそれぞれの相対的な確信度を推論するのである．

別の例として，禁止薬物の使用に関する検査を考えてみよう．検査対象者を無作為抽出し，血液検査によって違法な薬物を使用しているかどうかを推論したい．しかし残念なことに検査は完璧ではなく，ノイズが含まれてしまう．擬陽性や擬陰性が生じる可能性は低くない．さらに，この薬物を実際に使用している人は少ないことが事前に判明しており，この情報も考慮しなければならない．以上より，ありうる可能性は薬物使用の有無であり，これら 2 つの可能性は，薬物使用者の割合に関する既存の知識に基づく事前の確信度を持っている．薬物検査の結果データにはノイズが含まれるが，ベイズ推論によって 2 つの可能性に確信度を再分配できる．いずれ定量的に説明するが，たとえ薬物検査の結果が陽性でも，薬物使用に関する事後確率は驚くほど小さいことが少なくない．これは，薬物使用に関する事前確率がそもそも小さく，検査にノイズが含まれるためである．薬物検査に限らず，癌などの病気の診断においても同様である．現実におけるベイズ推論の応用例には，スパムメールの検出がある．スパムメールを自動的にフィルタリングするために，ベイズ推論によって新着メールがスパムであることの事後確率が計算されている．

要約すると，ベイズ推論の本質は，複数の可能性に確信度を再分配することである．確信度の分布は，初めは可能性に関する事前の知識を反映しているが，これは非常に漠然としている．新しくデータが観測されることで，確信度が再分配される．可能性がデータと一致した場合は，より多くの確信度が蓄積される．一方で，可能性がデータと一致しなければ確信度を失う．ベイジアン分析は，論理的に筋が通りかつ正確な方法で，確信度を再

分配する数学なのである．

2.2.　可能性は記述的モデルのパラメータ値である

ベイジアン分析において，確信度を分配する対象に対して可能性の候補を定義することは，重要なステップである．これが軽視できないステップである理由は，確信度の分配対象に含めなかった可能性は常に存在するからだ（例えば濡れた歩道の例では，宇宙人が巨大な涙を落としたのかもしれない）．とはいえ，自分たちの関心に含まれる範囲で可能性を選べばよい．分析後に，対象とした可能性の中でもっとも確信できるものがデータをよく説明しているかどうかを検証することができるからだ．もしデータがよく説明されていなければ，より広い範囲から可能性を選ぶことになるだろう．事後予測チェックと呼ばれるこのプロセスについては後述する．

血圧の薬に関する例をもう一度見てみよう．この例では，ある群の検査対象者には薬を投与した後で，別の群の検査対象者にはプラセボを与えた後で血圧を測定した．2群の血圧の傾向はどれほど違うだろうか？　すなわち，各群の代表的な血圧はどの程度違うだろうか？　群間の違いに関して，どの程度確信を持てるだろうか？　差の程度はデータの記述である．私たちが知りたいのは，どの記述が相対的に確信できるのか，あるいは確信できないか，である．

一般的にデータ分析は，データの記述の候補を集めることから始まる．この記述とは，データの傾向や広がりを表す数式である．式自体が持つパラメータ値と呼ばれる数値により，数式の正確な形が決定する．パラメータとは，データを生成する数学的な装置の調節つまみである．音楽プレーヤーのボリュームを変えたら，音の強さも変わるだろう，それと同じように，もしパラメータ値を変更すれば，生成されるデータの傾向も変わるのである．

統計学や数学における先行研究を読むと，いわゆる正規分布と呼ばれる釣り鐘型をした分布をよく目にするだろう．上述したボール製造の例（図2.3）でも触れているように，正規分布はデータの記述によく用いられる．正規分布は，平均と標準偏差という2つのパラメータを持つ．平均は分布の中心位置を決定するための，数式における調節つまみである．平均は位置パラメータと呼ばれることもある．標準偏差は，分布の幅やばらつきを決定するための，数式における別の調節つまみである．標準偏差は，スケールパラメータと呼ばれることもある．正規分布に関する数式は，パラメータ値を，データ値の確率に関する釣り鐘の形に変換する．

図2.4では，データに記述の候補となる正規分布を重ね合わせている．データはヒストグラムとして表されており，縦軸は各棒の幅となる範囲にどれだけのデータが含まれるかを表している．ヒストグラムはおおよそ単峰性で左右対称であるように見える．上段には，データの記述の候補として，平均が10で標準偏差が5である正規分布を重ね合わ

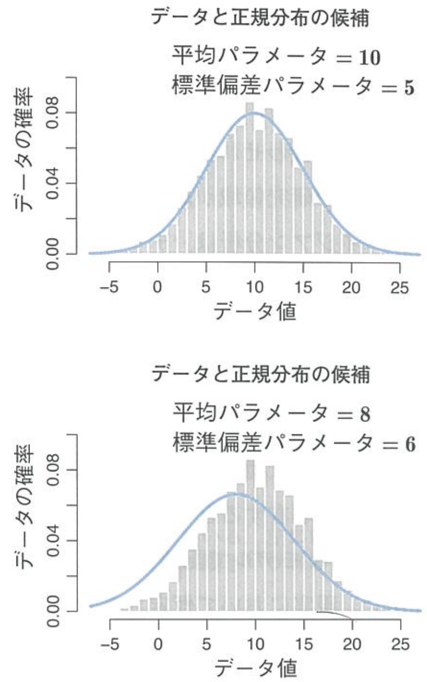

図2.4　2つのグラフは，同一のデータのヒストグラムである．ただし，候補となる正規分布の記述が異なる．ベイジアン分析では，候補となるパラメータ値の相対的な確信度を計算する．

せてある．このパラメータ値の選択は，データをうまく記述しているように思われる．下段には，平均が8で標準偏差が6という別のパラメータ値を持つ正規分布を重ねている．このデータの記述も説得力があるように見えるが，上段の例ほどではない．ベイズ推論の役割は，候補となるパラメータ値の事前確率を考慮しながら，相対的な確信度を正確に計算することにある．

現実的な応用例では，候補となるパラメータ値は，少数の離散的な値ではなく，無限の連続量の形をとることもあるだろう．正規分布の位置パラメータは，負の無限大から正の無限大までいかなる値もとりうる．ベイズ推論は，このような無限の連続量でも問題なく扱うことができる．

データを数学的に記述する際に，必須事項が2つある．第一に，数式は意味のあるパラメータを用いて，理解できるようにつくるべきである．知らない言語でデータを記述しても役に立たないように，解釈できないようなパラメータを用いて理解できない数式をつくっても意味がない．正規分布の例では，平均と標準偏差のパラメータは，それぞれ分布の位置とスケールを示しているため，意味が直接わかる．本書では一貫して，意味のあるパラメータを用いた数学的記述を心がける．すなわちベイジアン分析では，選ばれたモデルにより定義された意味のあるパラメータ値の候補に対して，確信度を再分配していくのである．

数学的記述における必須事項の2つ目は，記述的妥当性である．要するに，数式はデータの「ように見える」べきである．データの傾向とモデルの形式の間に，重要な乖離や体

系的な乖離があってはならない．明確な乖離が存在するとき，それが重要なものや体系的なものであるかを決定するのは，確定されたプロセスではない．研究の初期段階では，データを大雑把に，「ほどほど」に記述するだけでも満足できるかもしれない．そのようなやり方でも，既存の知識に照らし合わせれば興味深く新規な，意味のある傾向が見つかるかもしれないからだ．しかしその研究領域が成熟するにつれ，データをより正確に記述する必要が出てくるだろう．ベイジアン分析は，データの記述候補の相対的な確信度を検証する上で，非常に有用である．

データを数学的に記述することと，データの因果関係を説明することの違いを理解しておくことも重要である．図 2.4 のデータは，平均が 10 で標準偏差が 5 の正規分布でうまく記述されている．しかし，データがこのような形状になった原因まではわからない．パラメータは正規分布で定義された既知の数式においてのみ「意味がある」．パラメータ値は必ずしも実際の原因に関する意味を有するとは限らない．応用的な研究では，データを生成する現実のプロセスを数学的なモデルで記述したい場合もあるだろう．こういった場合には，パラメータや数式で，現実において仮定された状態やプロセスを表現してもよい．ボール製造の例（図 2.3）では，候補となるパラメータ値は工場で設定された「サイズ」として解釈できる．そして，背後にあるサイズにランダムな膨張が加わり，観測されるデータが定まるのである．しかし，単にデータの傾向を記述するだけであれば，物理的な状態やプロセスに言及する必要はない．本書では，様々な領域に広く応用可能で，直観的に理解可能なモデルを用いて，一般的なデータを記述することにする．

2.3. ベイジアン分析のステップ

一般的に，ベイジアン分析は以下のステップで行われる．

1. リサーチクエスチョンと関連するデータを特定しよう．データはどのようなスケールで測定されているだろうか？　どのデータ変数が予測される変数で，どのデータ変数が予測する変数だろうか？
2. 関連するデータを記述するモデルを定義しよう．数式とそのパラメータは，分析の理論的な目的に対して，意味があり適切なものでなければならない．
3. パラメータの事前分布を特定しよう．事前分布は，分析結果の読み手（例えば懐疑的な科学者）にも納得されるものでなければならない．
4. ベイズ推論を用いて，パラメータ値に確信度を再分配しよう．理論的に意味のある目的に照らし合わせて事後分布を解釈しよう（次の手順で示すように，モデルはデータを適切に記述していることを前提としている）．
5. 事後予測が十分正確にデータを模倣できているかを確認しよう（つまり，「事後予測チェック」を行う）．もし似ていなければ，記述モデルが異なっている可能性を考え

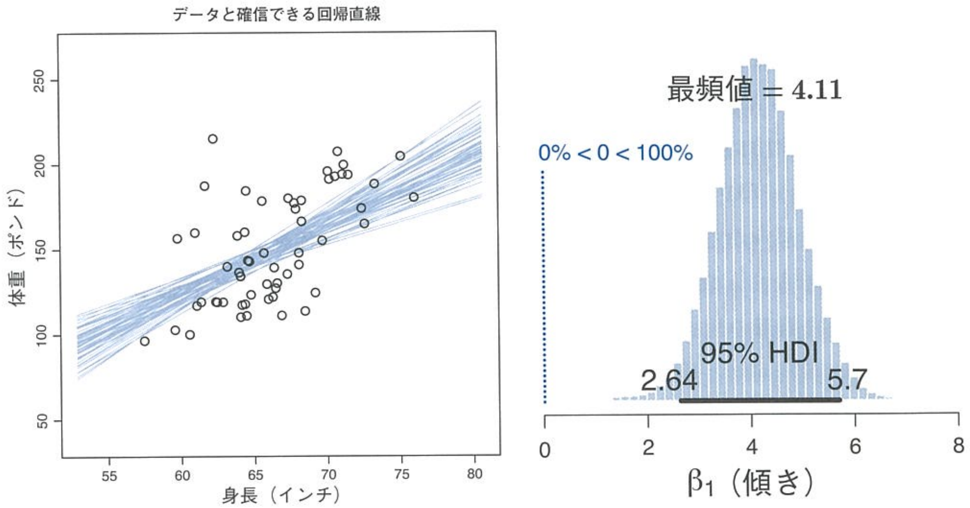

図 2.5　左の散布図では，データを円で表している．左には，事後分布から求められた確信できる回帰直線がいくつか重ね描かれている．右は，傾きパラメータ（式 2.1 における β_1）の事後分布を表している[d]．

よう．

おそらくこれらのステップを説明するためにもっとも良い方法は，現実的なベイジアン分析の例をあげることだろう．本章では導入が目的なので，以下の例では考察や専門的な細部は省略する．体重と身長の関係が知りたいと仮定しよう．日々の経験から，身長が高い人ほど体重が重いと考えるだろう．しかしここでは，身長が高くなるにつれてどの程度体重が増加するかや，その増加の程度の確からしさを知りたいとしよう．特に，身長から体重を予測できるかどうかに興味があるとしよう．

ステップ 1 では，関連するデータを特定するのであった．関心のある母集団から無作為に抽出された，成人 57 人の身長と体重のデータが手に入ったとしよう．身長はインチ単位で，体重はポンド単位で，いずれも連続量として測定されている．ここで身長から体重を予測してみたい．図 2.5 に散布図が示されている．

ステップ 2 では，研究上の興味において意味があるように，データを記述するモデルを定義する．この時点では，体重と身長の間におおよそどのような関係があるかがわかればよい．少なくとも成人全体の傾向としては，体重は身長に比例すると考えてよいだろう．よって体重は，身長の倍数にベースラインを加算することで予測できる，と記述しよう．予測される体重を $\hat{y}$（「y ハット」と読む）で，身長を x で表すことにする．すると，体重は身長の倍数にベースラインを加算することで予測できるというアイデアは，数学的には以下のように記述できる．

d）訳注：身長 55 インチは約 139.7 cm，80 インチは約 203.2 cm，50 ポンドは約 22.68 kg，250 ポンドは約 113.40 kg．

$$\hat{y} = \beta_1 x + \beta_0 \tag{2.1}$$

係数 β_1（ギリシャ文字の「ベータ」）は，身長が1インチ上昇したときに，予測される体重がどの程度増加するかを示している[2)]．式2.1ではベースラインは β_0 と表されており，身長が0インチの人の体重を意味している．ベースラインの値は，アプリオリに0であるべきだと思うかもしれないが，成人の体重と身長の関係を記述する場合にはその必要はない．なぜなら，成人の身長は0よりはるかに大きい一定の範囲に収まるはずだからだ．式2.1は線形の形式をとっており，β_1 が傾きを，β_0 が切片を表している．このようなモデルは線形回帰と呼ばれることが多い．

このモデルはまだ完璧ではない．予測される体重の周辺で，実際の体重がランダムに変動することを記述しなければならない．簡単にするために慣例的な正規分布を用いて（詳細な説明はセクション4.3.2.2に譲る），実際の体重 y は予測値 $\hat{y}$ の周辺で正規分布に従うようにランダムに分布し，標準偏差は σ（ギリシャ文字の「シグマ」）となると仮定しよう．この関係は以下のように表すことができる．

$$y \sim \mathrm{normal}(\hat{y}, \sigma) \tag{2.2}$$

記号 "~" は「のように分布する」ことを意味している．式2.2は，$\hat{y}$ に近い y の値が生じる確率がもっとも高く，$\hat{y}$ よりも高いまたは低い y の値が生じる確率は相対的に低いことを意味している．$\hat{y}$ から離れるにつれて確率がどのように減少するかは，標準偏差 σ のもとで正規分布がどのような広がり方をしているかで決定される．

式2.1と式2.2を統合した完全なモデルは，合計で3つのパラメータを持っている．それは，傾き β_1，切片 β_0，「ノイズ」すなわち標準偏差 σ である．これらはいずれも意味があるパラメータである．具体的には，傾きのパラメータは，身長が1インチ伸びた場合に体重がどの程度増加するかを示している．標準偏差のパラメータは，予測される値の周辺で体重がどの程度変動するかを表している．この種の線形回帰と呼ばれるモデルは，第15章，第17章，第18章で詳細に説明する．

分析におけるステップ3は，パラメータの事前分布を特定することである．対象の母集団の体重や身長に関する，先行研究や公的に立証可能な知見を，事前分布に用いることができるかもしれない．または，社会的な相互作用の中で共通認識となった経験に基づいて，事前分布を控えめに設定することもできるかもしれない．しかしこの例では，傾きと切片のとりうる範囲に，確信度が0を中心としてほぼ均一に分布した，漠然とした事前分布を使用する．ノイズ（標準偏差）のパラメータに関しても，0から非常に大きい値まで一様に分布する曖昧な事前分布を使用する．このような事前分布を選ぶということは，

2) β_1 が，x が1単位上昇したときの $\hat{y}$ の増加程度を表していることを証明しよう．まず，身長が x のとき，予測される体重は $\hat{y}_x = \beta_1 x + \beta_0$ である．次に，身長が $x+1$ のとき，予測される体重は $\hat{y}_{x+1} = \beta_1(x+1) + \beta_0 = \beta_1 x + \beta_1 + \beta_0$ である．つまり，予測される体重の変化は $\hat{y}_{x+1} - \hat{y}_x = \beta_1$ である．

事後分布に与える影響にはほとんどバイアスが存在しないことを示唆する．

ステップ4では事後分布を解釈する．ベイズ推論は漠然とした事前分布からデータに合致する値まで，様々なパラメータ値に確信度を再分配する．事後分布は，データを所与としたときの，確信できる β_0，β_1，σ の組み合わせを意味している．図2.5（右）は傾きパラメータ β_1 の事後分布を示している（なお他の2つのパラメータは割愛した）．図2.5はデータの分布ではなく，パラメータ値の分布であることに注意してほしい．図2.5の青い棒は，連続的な傾きの値の候補に，それぞれどの程度の確信度があるかを示している．シャーロック・ホームズの例や無罪放免の例，離散的な平均値の候補の例（図2.1〜図2.3）における青い棒と似ていることがわかるだろう．図2.5の事後分布によれば，もっとも確信できる傾きの値は約4.1である．つまり，身長が1インチ伸びるにつれて，体重は4.1ポンド上昇する[e)]．事後分布は，連続的に推移する傾きの値に関する相対的な確信度も表しているため，推定された傾きの不確かさも判明する．不確かさを要約する方法の一つは，もっとも確信でき，かつ分布の95% をカバーするような値の範囲を見つけることである．この範囲は最高密度区間 (highest density interval; HDI) と呼ばれ，図2.5の分布では底面に黒い横線で示している．95%HDIの範囲内の値は，HDIの範囲外の値よりも確信できる（つまり，より高い確率「密度」を持っている）．また，HDIの範囲内の値が得られる確率は95% である．57個のデータ点を所与としたとき，傾きの95%HDIは1インチあたり2.6ポンドから5.7ポンドまでの範囲をとる．もっとデータが増えれば，傾きの推定はより正確になるだろう．つまり，HDIはより狭くなる．

図2.5は，傾きが0の場合は事後分布のどこに位置するかも示している．この例では，0はいかなる確信できる傾きの値とも大きく離れている．よって傾き0は，身長と体重の関係に関する記述としては「棄却」してよいだろう．しかし，このような0の位置関係に関する離散的な決定は，完全な事後分布を生成するベイジアン分析そのものではない．

読者の多くは，すでに帰無仮説検定 (NHST) について学んだことがあると思う．帰無仮説検定では，t 値のような要約統計量の標本分布 (sampling distributions) を用いて，p 値が計算される（もしこれらの用語を知らなくても，帰無仮説検定については第11章で論じるので心配には及ばない）．図2.5における事後分布は，標本分布ではなく，p 値とは関係がないことに注意してほしい．

事後分布を理解する上では，確信できる回帰直線の例を，散布図上にプロットしてみることも有効である．図2.5（左）には，事後分布から求めたランダムな確信できる回帰直線をいくつか示している．それぞれの直線は確信できる β_1 と β_0 の組み合わせにより，$\hat{y} = \beta_1 x + \beta_0$ をプロットしている．「最良の」直線を1本引くのではなく，データを所与としたときに確信できる可能性の範囲を示している．

ステップ5では，モデルのもっとも確信できるパラメータ値が，実際のデータをよく

e) 訳注：1インチは約2.54 cm，4.1ポンドは約1.86 kg なので，このデータでは1 cm あたり732 g 上昇している．

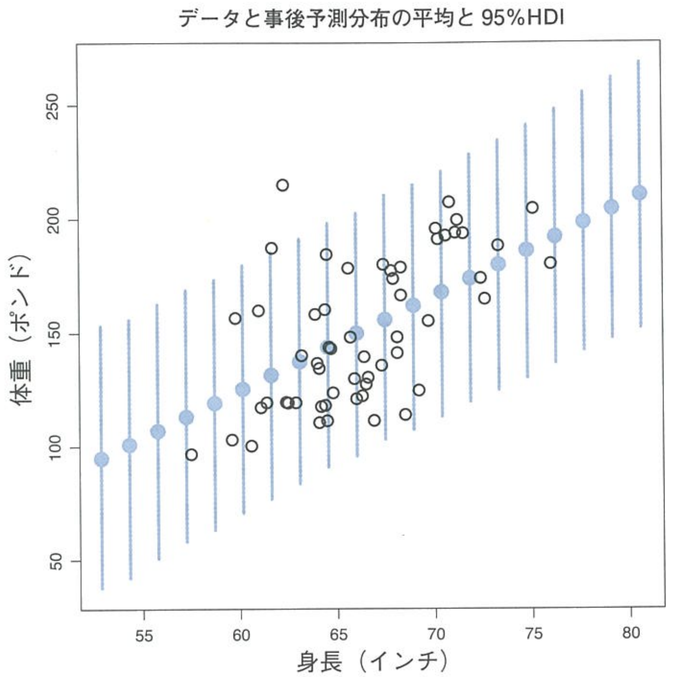

図 2.6　図 2.5 のデータを，任意の身長における予測される体重の事後分布とともに示した．垂直方向の棒はそれぞれ，予測される体重に関する 95% 確信できる範囲を表している．各棒の中央にある点は，予測される体重の平均値を示している．

模倣できているかという観点から，モデルをチェックする．これは「事後予測チェック」と呼ばれる手続きである．モデルの予測が，体系的かつ意味があるほどにデータから逸脱しているかどうかを判断するための，唯一絶対の方法は存在しない．なぜなら，体系的な逸脱には無数の定義が存在するからである．一つのやり方は，モデルから予測されるデータの要約と，実際のデータをプロットすることである．パラメータ β_1，β_0，σ の確信できる値を式 2.1 や式 2.2 に投入し，任意の x（身長）における y（体重）をランダムに生成してみる．この手続きを何度も何度も繰り返して，モデルのもとで得られるデータの代表的な分布を生成する．このシミュレーションの結果は図 2.6 の通りだ．予測された体重の値は，もっとも確信できる 95% の範囲として垂直の棒で要約されている．各棒の中央にある点は，予測された体重の平均値を表す．グラフを一目見て，実際のデータは予測されたデータでよく記述されていることがわかるだろう．実際のデータは，モデルにより予測された傾向や区間から，体系的に逸脱してはいないようにみえる．

もし実際のデータが，予測されたデータの形状から体系的に逸脱しているように見えたら，他の記述モデルを考えてみよう．例えば，実際のデータは非線形の傾向を示すかもしれない．その場合は，非線形の傾向に対処できるように，モデルを修正すればよい．ベイズ用ソフトウェアを用いれば簡単に，そのような修正や，非線形の傾向を記述するパラメータの推定ができる．さらに，データの分布に関する特徴を検証することもできる．例えば，もし正規分布から逸脱しているような外れ値が存在するとしたら，裾が長い分布を

用いるようにモデルを修正すればよい．これもベイズ用ソフトウェアを用いれば簡単にできる．

ここまで非常に現実的な例を用いて，ベイジアン分析を行うための5つのステップを紹介してきた．本書では，様々な応用的事例や記述モデルに対して，どのように同様の分析を行えばよいかを説明する．2 群の比較におけるベイジアン分析については，Kruschke (2013a) の論文で簡潔かつ詳細に紹介されている．この論文では，古典的な t 検定の抱える問題についても説明されている．重回帰分析に対するベイジアン分析の適用例は，Kruschke, Aguinis, and Joo (2012) の論文で紹介されている．事後予測チェックの展望については，Kruschke (2013b) の論文や本書の（特に）セクション 17.5.1 を参照してほしい．

2.3.1．計量的なモデルを必要としないデータ分析は可能か？

これまで概説してきたように，ベイジアン分析は意味があるようにパラメータ化された記述モデルに基づいている．そのようなモデルが使えない，あるいは必要ない状況は存在するのだろうか？

パラメータ化されたモデルが用いられないであろう状況の一つは，いわゆるノンパラメトリックなモデルである．しかし実際には，ノンパラメトリックなモデルにもパラメータは存在するため，この命名は誤解を招きかねない．ノンパラメトリックなモデルは，潜在的には無限のパラメータを持っているのだ．犬の体重を記述するという，単純な例で説明しよう．あらゆる犬種から多くの犬を無作為に抽出し，体重を量ってみる．おそらく，体重の分布は単峰性にはならないだろう．むしろ，犬種ごとにいくつかの体重の下位クラスターができると予想される．もちろん中には，体重の分布が似通う犬種もあれば，犬種が特定できないような犬も多く存在するだろう．そして犬のデータを集めるほど，以前のデータには含まれていなかったような，新たな下位クラスターに属するデータが得られることもあるだろう．つまり，記述モデルに含めるべきクラスターの数に関して，明確な基準はないのだ．むしろデータに基づいて，異なるクラスタリングどうしの相対的な確信度を推論しよう．各クラスターは自前のパラメータを持っているため（例えば位置やスケールのパラメータ），データが無限に存在するとき，モデルにおける推論されるパラメータの数も無限になる．無限にパラメータ化されたモデルには，他にも様々な種類がある．ベイズ的でノンパラメトリックなモデルに関するチュートリアルは，Gershman and Blei (2012) を参照してほしい．最近のレビュー論文は Müller and Mitra (2013) を，教科書としては Gelman et al. (2013) を参照してほしい．本書ではベイズ的なノンパラメトリックモデルは扱わないことにする．

一見しただけではパラメータ化されたモデルが適用できないように思える状況はいくつもある．例えば，ある珍しい病気の診断結果が陽性であったときに，実際にその病気に罹患している確率が知りたいとしよう．この場合は，パラメータは連続的な分布ではな

く，離散的な状態を指している．しかしベイジアン分析は，このような状況にも適用できる．病気の診断の例では，パラメータは背後にある個人の健康状態であり，「罹患している」または「罹患していない」のいずれかの値をとる．ベイジアン分析では，観察された検査結果に基づいて，このような2値のパラメータ値に対して確信度を再分配する．検査の結果得られるものが，完璧に決定的な情報ではなく確率的な情報である点を除けば，シャーロック・ホームズが離散的な可能性について考えている図2.1の例とそっくりである．このような状況におけるベイズの計算については，第5章で説明する（特に，表5.4を参照）．

最後に，分析者がいかなるパラメータ化されたモデルも使いたくない場合があるかもしれない．たとえ，とてつもなく柔軟で無限にパラメータ化されたモデルであったとしても，である．このような場合には，ベイズ的手法は適用できない．もっとも，数学的なモデルは驚くほど便利なツールであるので，このような状況は稀である．モデルを用いずにデータから推論を行うケースの1つは，帰無仮説検定におけるリサンプリングやブートストラッピングと呼ばれる方法である．これらの方法では p 値を計算して判断を行うのだが，第11章で論じるように p 値には根本的かつ論理的な問題がある．これらの方法では，データの特徴に関する確からしさの程度について，非常に限られた表現しかできない．一方でベイズ的手法では，不確かさの情報を前面に押し出すことができる．

2.4. エクササイズ

以下のエクササイズは一例である．すべてのエクササイズに取り組みたい場合は，https://sites.google.com/site/doingbayesiandataanalysis/へアクセスしてもらいたい．

エクササイズ 2.1. [目的：確率の数学的モデルを積極的に操れるようになること]

ボードゲームで4面ダイスを振る状況を想像してみよう．四面体のダイスでは，それぞれの面は正三角形で構成されている．ダイスを振ると，ある面が下になって，ピラミッドのように他の3面を見ることができる．各面には1から4までのいずれかの数字が割り振られており，底面の数字は他の面の底辺に印字されている点の集合を見ればわかるようになっている．底面の数字を x と表すことにしよう．x の確率に関して，以下3つの数学的な記述を考えてみよう．モデルA：$p(x) = 1/4$，モデルB：$p(x) = x/10$，モデルC：$p(x) = 12/(25x)$．それぞれのモデルで，x の値を変化させて $p(x)$ を求めてみよう．それぞれのモデルで，どのようなバイアスが存在する（または存在しない）ことが表現されているか，言葉で表してみよう．

エクササイズ 2.2. [目的：データがどのように確信度を変化させるかに関して，積極的に考えられるようになる]

前のエクササイズで紹介した四面体のダイスと，このダイスの確率に関する3つの候補を思い出そう．はじめは，ダイスの何を信じればよいのか見当もつかないだろう．ダイスは公平で，どの面が下になる確率も等しいかもしれない．ダイスにはバイアスがあって，数字が大きい面ほど下になりやすいかもしれない（数字を表す点は重い宝石で作られていて，より点が多い面ほど下になりやすいかもしれないからだ）．あるいはまた，数字が大きい面ほど下になりにくいかもしれない（点は弾力性のあるゴムで作られているかもしれないし，点は突起になっているかもしれないからだ）．よって最初のうちは，3つのモデルに関する信念は，$p(A) = p(B) = p(C) = 1/3$ と記述できるだろう．さて，このダイスを100回振って，1の目が出た回数が25回，2の目が出た回数が25回，3の目が出た回数が25回，4の目が出た回数が25回，という結果が得られたとしよう．このようなデータが得られたとき，モデルに対する信念は変わるだろうか？　どのモデルがもっともありえそうだろうか？　あるいは，このダイスを100回振って，1の目が出た回数が48回，2の目が出た回数が24回，3の目が出た回数が16回，4の目が出た回数が12回，という結果が得られたとしよう．さあ，今度はどのモデルがもっともありえそうだろうか？

存しておいたほうがよいだろう）．

エクササイズ 3.3. [目的：Rのコマンドシンタックスに詳しくなる]

SimpleGraph.R のプログラムを実行し，($y = x^3$) という 3 次関数が $x \in [-3, +3]$ の範囲上に描画されているのを確認しよう．出力されたグラフを任意のフォーマットで保存しよう．また，作図のためのコードにコメントをつけて，グラフと一緒に保存しておこう．

第4章

確率と呼ばれるものはいかなるものか？

Contents

Oh darlin' you change from one day to the next,
（おお，ダーリン，あなたは次から次へと変わっていくのね）
I'm feelin' deranged and just plain ol' perplexed.
（わたしは狂ったり，困惑したり）
I've learned to put up with your raves and your rants:
（わたしはあなたの絶賛と暴言を我慢し学んできたわ）
The mean I can handle but not variance.
（わたしが扱えるのは平均ではなく分散なのね）[1]

推測統計的な手法は，可能性に対して私たちが持つ不確実性を正確に測定する．不確実

1) 本章では，確率分布の考え方について説明する．確率分布の考え方には，平均と分布の分散の技術的な定義がある．このポエムは，口語遊びである．

性は確率によって測定される．したがって，それについて推論する前に，私たちは確率の特性をしっかり把握しなければならない．本章では，確率の基本的な考え方を紹介する．あなたにとって，本章が短すぎると思われる場合は，本章のトピックについて書かれたよい参考書 Albert and Rossman (2001, pp. 227-320) がある．

4.1．すべての可能な出来事のセット

私はいま，コインを投げよう（フリップしよう）としている．表 (head) が出るのはどのくらいだろうか．裏 (tail) が出るのはどのくらいだろうか[2]．真ん中 (胴体 torso) が出るのはどのくらいだろうか．私たちがそれぞれの結果の可能性をじっくり考えるとき，起こりうるすべての結果を1つのまとまりとして考えていることに注意してほしい．真ん中が出ることは，起こりうる結果の1つではない．また，コインを1回指で投げることで出るのは1回の結果でしかないことにも注意してもらいたい．1回のコイン投げで表と裏の両方が出るはずがない．この結果は相互に排他的である．

私たちが，結果がどのくらい起こりうるかについて尋ねるときは，いつも起こりうる結果のセットを念頭に置いている．このセットは，すべての起こりうる可能性を網羅しているし，結果はすべて相互に排他的である．このセットは**標本空間** (sample space) と呼ばれている．標本空間は，私たちが世界を観測するために用いられる測定活動によって決定される．本書を通じて紹介されるすべての応用例について，測定を行うのにきっちりと定義された操作があることを，私たちは当然のこととして受け止めている．例えば，コイン投げについて，コインを投げ，それをキャッチするために明確に定義された方法が当然あるはずである．つまりコインが動かなくなり，どちらの結果になったのかしっかり宣言できるほど十分安定的な状態になったときに，私たちはしっかりと判定を下すことができる[3]．あるいは別の例をあげよう．人物の身長を測定する際には，私たちは身長計に対してとるべきポーズについて，しっかり定義されていることを当然のこととして受け止めているだろう．その上で目盛りを落ち着いて読み取り，ある人の身長に特定の値を宣言することができるのである．機械的な操作化，数学的形式化，および測定の哲学的研究については，それぞれについて1冊全体をかけて論じている本がいくつもある．ここではこの1段落で，それを終わらせてしまおう．

コインを投げるとき，コインの表が出る可能性を考えてみよう．もしコインが公平なら

2） 政府によって製造される多くのコインは，一方の側に重要人物の顔の絵がある．この側面は「頭 (heads)」または技術的に「表 (obverse)」と呼ばれる．裏面はめったに描かれることはないが，「頭」の反対である「尻尾 (tails)」と呼ばれている．

3） 実際には，投げられた（弾かれた）コインはいつも 50% の確率で表が出るが，スピンしたコインは表／裏が出る確率が平等でないことが知られている (Gelman & Nolan, 2002)．投げることとスピンさせることの違いが重要であると思うなら，本書でコインを投げたところを「スピンした」と書いてあれば，実際は「フリップした（弾いた）」のだな，とあなたの心で置き換えておいてほしい．コイン投げ確率の実証的かつ理論的研究については 例えば， Diaconis, Holmes, and Montgomery (2007) を参照．

ば，投げたコインの表が出る確率は約 50% にならなければならない．もしコイン（あるいは投げる手順）に偏りがあるならば，投げたコインの表が出る確率は 50% 以上，あるいは 50% 以下になる傾向があるだろう．コインの表が出る確率は，パラメータラベルの θ（ギリシャ文字で「シータ」）で表すことが可能である．例えば，$\theta = 0.5$（「シータが 0.5 と等しい」）のとき，コインは公平である．

また，私たちは，コインが公平であるという確信の程度を考慮することが可能である．私たちは，コインが政府の造幣局で製造されたことを知っていたなら，コインは公平であるという高い確信を持つ．あるいはコインが Acme Magic and Novelty 社[a]で製造されたことを知っていたなら，そのコインは偏りがあるという高い確信を持つ．パラメータの確信の度合いは $p(\theta)$ で表すことが可能である．もしコインが政府によって発行されたものである場合，このコインは公平であるという強い確信を持つだろう．例えば，私たちは $p(\theta=0.5) = 0.99$，つまり「θ が 0.5 である確率が 99% である」と信じているといえるだろう．もしコインが Novelty 社によって製造された場合，私たちは，このコインは偏りがあるというつよい確信を持つかもしれない．例えば，$p(\theta = 0.5) = 0.01$ または $p(\theta=0.9) = 0.99$ であると信じている，というように．

コインの表と裏が出る両方の「確率」とバイアスによる「確信の度合い（程度）」はいずれも標本空間を参照して得られる．コイン投げのための標本空間は，2つの可能な結果（表と裏）で構成されている．コインのバイアスのための標本空間は，とりうる値の連続値で構成されている．すなわち $\theta = 0.0, \theta = 0.01, \theta = 0.02, \theta = 0.03$, および $\theta = 1.0$ までの間のすべての値である．与えられたコインを投げるとき，私たちは，表か裏の標本空間からサンプリングする．異なるバイアスを持っている可能性のあるコインの袋からランダムにコインをつかむとき，私たちは可能なバイアスの標本空間からサンプリングしているのである．

4.1.1．コイン投げ：あなたが気にしなければならないのはなぜか

コインの公平さは，賭け金の高いゲームにおいては重要かもしれない．しかし，コインを投げることとその結果を気にすることは，人生においてそれほど多くない．ではなぜ，わざわざコイン投げの統計について調べるのか？

なぜなら，コイン投げは私たちが関心を示している他の無数の現実の事象の代わりとなるものだからである．ある心臓手術において，患者が1年以上生存しているかどうかを患者の結果として分類するとしたら，その患者が1年以上生存する確率はどれくらいか知りたいと思うだろう．ある薬物について，頭痛があるかどうかを結果として分類するかもしれないし，分類すれば頭痛になる確率を知りたいと思うだろう．質問紙調査の場合，同意するかどうかの程度が結果であり，それぞれの反応確率について知りたいのである．

a） 訳注：米国の漫画に出てくる架空の会社．この会社の詳細については各自で検索してみてほしい．

2人の候補がいる選挙においては，結果は候補Aまた候補Bのどちらかが当選する，である．そして，選挙の前に，世論調査から候補Aが勝つ確率を推定したい．あるいはまた，項目の結果が正しいか誤りのどちらかであり，多項目試験の正確性を測定することで算数能力の研究をしているのかもしれない．はたまた，あなたは脳の研究をしていて，異なる個体集団における特定の認知プロセスの脳の側性化を研究しており，左側か右側のいずれが反応するかを結果として，この個体集団において左側性化がみられる確率を推定したいとしているかもしれない．

それそのものは魅力的でないかもしれないが，私たちがコイン投げについて議論しているときはいつでも，あなたが実際に関心を持っている領域に当てはめることができることを心に留めておいてほしい．コインは似た用途の一般的な代表例にすぎない．

4.2. 確率：頭の外側か内側か

私たちはときどき，世界の「向こう側」にある結果の確率について話をする．投げられたコインの面は結果である．私たちはコイン投げを観測することができ，表が出る確率はいくつかのコイン投げを観測することで推定することができる．

しかし，はっきりと「向こう側」にあるとはいえない確率の話をしたり，むしろただ「頭の内側」について持ちうる信念の話をする．コインの公平さについての確信は，頭の内側の例である．コインは，固有の物理的なバイアスがあるかもしれないが，いまはそのバイアスに対する私たちの信念についての話である．私たちの信念は，相互に排他的かつ網羅的な可能性の空間を参照している．私たちがコインの入った袋からランダムにサンプリングするように，私たちの信念からランダムにサンプリングしているということは，奇妙に思えるかもしれない．それでも私たちが示すように，頭の外側の確率の数学的特性と頭の内側に対する信念は本質において同じである．

4.2.1. 頭の外側で：長期の相対度数

頭の外側の事象について，それぞれの起こりうる確率の長期的相対度数である確率について考えるのは直観的である．例えば，公平なコインにとって表の確率は0.5であるという場合，それが意味するのは，コインを何度も投げたとき，表が出るのは約50%だろうということである．長期的に，コインを何度も何度も投げた後，表の相対度数は限りなく0.5に近づく．

長期的相対度数を見つけ出すのは，2つの異なる方法によって可能である．第一に，実際に標本空間から多数回サンプリングし，各事象が発生した回数を集計することにより，それを近似する方法である．第二に，数学的にそれを導き出すことである．これらの2つの方法は，これから順に探っていく．

4.2.1.1. 長期的相対度数をシミュレートする

公平なコインから表が得られる長期相対頻度を知りたいとする．コイン投げを長々と反復していくと約50%の表が得られなければならないというのは，明白であると思えるかもしれない．しかし，それがそれほど明らかではないということにしておこう．つまり，私たちが知っているのは，“表”または“裏”を生成するというある程度基本的なプロセスがある，ということだけである．そのプロセスには，値が$\theta = 0.5$であるθと呼ばれるパラメータを持っているとしよう．それが私たちの知っているすべてなら，繰り返しのプロセスからサンプリングするだけで“表”が出る長期確率を近似することができる．プロセスからN回サンプリングし，“表”が出現した回数を集計し，相対頻度によって表の確率を推定する．

コイン投げのように手作業でプロセスをサンプリングするのは，面倒で時間がかかる．その代わりに，コンピュータにはるかに速く繰り返しサンプリングをさせることが可能である（そして，うまくいけばコンピュータの作業は手作業より退屈しなくて済む）．図4.1は，公平なコインを何度も投げたコンピュータのシミュレート結果を示している．Rプログラミング言語には，これから頻繁に使用することになるが，組み込まれた擬似乱数発生器がある[4]．1回目のコイン投げでは，コンピュータがランダムに表か裏を生成する．その後，それまでに得られた表の割合を計算していく．もし1回目のコイン投げで表だった場合，表の割合は$1/1 = 1.0$である．もし，1回目のコイン投げで裏だった場合，表の割合は$0/1 = 0.0$である．そして，コンピュータは，ランダムに2回目の表や裏を生成し，これまでに得られた表の割合を計算する．順序配列が表表である場合には，表の割合は，$2/2 =1.0$である．順序配列が表裏または裏表である場合には，表の割合は，$1/2 = 0.5$である．順序配列が裏裏である場合，表の割合は，$0/2 = 0.0$である．そして，コンピュータは，3回目の表または裏を生成し，すみやかにこれまで投げた中から表の割合を計算する．図4.1は，表が出てくる割合の推移を連続的に示している．

図4.1より，長い配列の最後には表の割合が0.5に近いが，正確に等しく0.5ではないことに注意してほしい．このずれは，長く続けたときでも，有限確率標本にしかならないし，事象の相対度数が事象の真の根底にある確率と一致するという保証はないことに気づかせてくれる．これが長期的な相対頻度確率に近似しているという理由である．

4.2.1.2. 長期相対頻度を導出する

ときどきではあるが，状況が十分に数学的に単純である場合，私たちは正確な長期的な相対頻度を導出することができる．公平なコインの場合は，そのような単純な状況の1つである．コインの標本空間は，コインの表と裏が起こりうる結果で構成されている．ま

4) 擬似乱数生成器 (PRNGs) は，実際にはランダムではない．つまり実際には決定論的である．しかしそれが生成する配列の性質は，ランダムなプロセスの特性を模倣している．本書で使用される方法は，PRNGs の品質に大きく依存しており，（例えば，Deng & Lin, 2000; Gentle, 2003）などで集中的に研究されている．

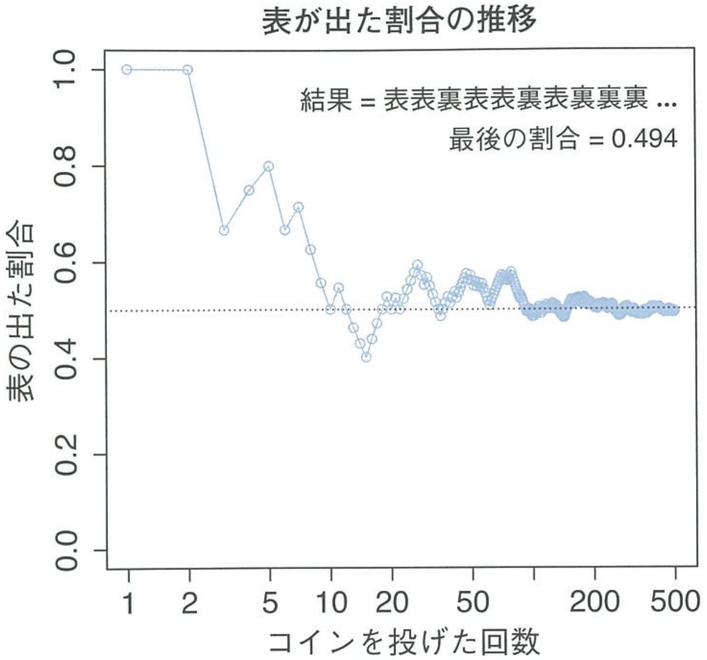

図 4.1 コインを投げているときに表の割合を実行している．最初の数回投げた結果の詳細だけでなく，多く投げた後の長期的な傾向を見ることができるように，x 軸は対数目盛りでプロットされている．この図を生成するための R コードはセクション 4.5 で説明されている．

た公平性を仮定することで，それぞれの結果が等しく起こりうることを知っている．そのため，表の長期相対頻度は，正確に 2 つのうち 1 つ，すなわち，1/2 でなければならない．また裏の長期相対頻度も正確に 1/2 にならなければならない．

この技術は，容易に他の単純な事態に拡張される．例えば，標準的な 6 面体ダイス（サイコロ）を考えてみる．これは，6 つの起こりうる結果，すなわち 1 の目，2 の目，...，6 の目を持っている．サイコロが公平であると仮定した場合，それぞれの結果の長期相対頻度は正確に 1/6 になるはずである．

サイコロの 6 面に異なる目を置くとしよう．具体的には，ある面に 1 の目，2 つの面に 2 の目，残りの 3 つの面に 3 の目を置いたとする．とはいえまだ 6 つの面のそれぞれが均等に起こりうることを前提としている．このとき，1 の目が出る長期相対頻度はちょうど 1/6 であり，2 の目の長期相対頻度はちょうど 2/6 であり，3 の目の長期相対頻度はちょうど 3/6 である．

4.2.2. 頭の内側で：主観的信念

米国政府によって発行されるコインが公平であることを，あなたはどれくらい強く信じているだろうか？　あなたはコインが正確に公平であるというよりもわずかに異なる可能性がある，と思っているのなら，どのくらい強く，表が出る確率が $\theta = 0.51$ または $\theta = 0.49$ であると信じているのだろうか？　代わりに，古くて非対称でいびつなコインを考えたとき，あなたはそれが本当に $\theta = 0.50$ であると信じられるだろうか？　マジックショップで購入したコインついてはどうだろうか？　ここでは，コインの表が出る本質的確率の真値についての話をしているのではない．私たちは，各確率の信念の度合いについて話をしているのだ．

主観的な信念を明示するために，それぞれの起こりうる結果をどのように考えるかについて明らかにする必要がある．柔軟で直観的な信念をはっきりさせることは，難しいだろう．次のセクションでは，主観的な信念を「調整」する 1 つの方法を模索し，続くセクションでは，数学的信念の度合いを記述するための方法について考える．

4.2.2.1. 好みによって主観的信念を調整する

旅行者に影響を与える可能性のある簡単な質問について考えてみる．次のお正月にインディアナポリス周辺の州間高速道路を閉鎖するような吹雪があるだろうと，あなたはどのくらい強く信じているだろうか？　あなたの仕事はこの質問に対し，正確にあなたの信念の確率を反映している 0 から 1 の間の数字を答えることだ．そのような数値を出す 1 つの方法は，他の確率がはっきりした事象と比較して，あなたの信念を調整することである．

比較する事象として，袋に入ったビー玉の試行を考えよう．袋に 10 個のビー玉を入れた．赤いビー玉 5 個，白いビー玉 5 個である．袋の中身を混ぜ，ランダムにビー玉を引いた．赤いビー玉が出る確率はもちろん $5/10 = 0.5$ である．お正月のインディアナポリスに雪が降ることを考慮するための比較として，ビー玉の袋の例を使用する．

次の 2 つのギャンブルについて考えてみよう．あなたは以下の 2 つから選択できる．

- ギャンブル A：次のお正月にインディアナポリスで吹雪による交通閉鎖がある場合，あなたは 100 ドルもらえる．
- ギャンブル B：赤いビー玉 5 個と白いビー玉 5 個が入っている袋から赤いビー玉を引いた場合，あなたは 100 ドルもらえる．

あなたはどちらのギャンブルを選ぶだろうか？　もしあなたがギャンブル B を選ぶなら，それはあなたがインディアナポリスでの吹雪による交通閉鎖が 50/50 よりも少ないと思っていることを意味する．つまり，少なくとも今あなたは吹雪による交通閉鎖の確率について，主観的信念が 0.5 未満であることを知っている．

さらに別のギャンブルを比較して考えることで，確信の度合いを絞り込むことができる．次の 2 つのギャンブルについて考えよう．

- ギャンブル A：次のお正月にインディアナポリスで吹雪による交通閉鎖がある場合，あなたは 100 ドルもらえる．
- ギャンブル C：赤いビー玉 1 個と白いビー玉 9 個が入っている袋から赤いビー玉を引いた場合，あなたは 100 ドルもらえる．

あなたはどちらのギャンブルを選ぶだろうか？　もしギャンブルAを選ぶなら，それはあなたがインディアナポリスでの吹雪による交通閉鎖が10%よりも多いと思っていることを意味する．上と合わせて考えると，2つのギャンブルを比較することが，あなたの主観的確率は0.1から0.5の間のどこかにあることを教えてくれる．この対立するギャンブルの選好を続けていくことで，あなたの主観的信念をより正確に調整していくことができる．

4.2.2.2.　数学的に主観的な信念を記述する

標本空間内にいくつかの起こりうる結果がある場合，すべての起こりうる結果についてあなたの主観的な信念を調整し続けると，手間がかかりすぎるかもしれない．その代わりに，あなたは自身の信念を集計する数学的関数を使用することが可能である．

例えば，あなたは，平均的な米国の女性の身長は1.64 mであると思っているかもしれないが，平均はその値よりいくらか上下に広がった確率であるだろう[b]．あなたの平均身長についての信念の度合いが，1.24 m，や1.27 m，や1.30 mについてはどうなっているのか，そこから1.85 m，1.88 m，や1.91 mではどうか，等々，特定していくのは不可能だろうし，あまりにも手間がかかりすぎる．だから，代わりに1.63 mでもっとも高くなり，左右対称に減衰していく釣鐘型曲線で信念の度合いを説明したらどうだろう．最良の主観的な信念を取り込むまでは，曲線の幅と中心を変更することができる．この本の後半で，私たちはこのような関数について正確な数式とともに話をするが，ポイントは，数学的関数として信念の度合いを記述するための曲線を定義することができる，という考えを理解することだけである．

4.2.3.　確率は起こりうることに数字を指定する

一般的に，確率は，それが頭の中にあるのか外にあるのかどうかは別にして，相互に排他的な確率のセットに数字を指定するだけの方法である．「確率」と呼ばれる数字が満たす必要がある特性は，次の3つだけである (Kolmogorov, 1956).

1. 確率値は非負でなければならない（すなわち，0または正）．
2. 全体の標本空間内におけるすべての事象の確率の合計は，1.0でなければならない（そうでなければ標本空間はあらゆる可能性を使い果たしていない，すなわち，空間内のいずれかの事象が発生しなければならない）．
3. 任意の2つの相互に排他的な事象について，どちらか一方が発生する確率は，個々の確率の和である．例えば，公平なサイコロで，3または4が出る確率は1/6 + 1/6 = 2/6である．

b）訳注：原文の単位はインチだが，この章だけは読者がイメージしやすいよう，メートル法に換算した．

ある出来事に与えられた数字が，これらの3つの特性を反映していれば，以下で論じる確率のすべての特性を持つことになる．つまり，確率を，世の中の結果の長期相対頻度としてみなすかどうか，また主観的信念の大きさとしてみなすかどうかによらず，数学的には同じように機能する．

4.3.　確率分布

確率分布は，単にすべての起こりうる結果とそれに対応する確率の一覧であると思ってほしい．コインであれば，その確率分布は明らかである．すなわち，私たちは2つの結果（裏と表）とその2つの対応する確率（θと$1-\theta$）の一覧が書ける．しかし，結果の他のセットでは，分布はより複雑になるだろう．例えば，ランダムに選択された人物の身長について考えるとしよう．身長が1.53 mである確率があり，また1.75 mである確率があり，その他同じように，ある高さについてあらゆる確率がある．身長のように，結果が連続的であれば，確率の考え方は後で見るように，いくつかの微妙な違いがある．

4.3.1.　離散分布：確率質量

標本空間が離散的な結果で構成されている場合，私たちは，それぞれ個別の結果の確率について話すことができる．例えば，投げられたコインの標本空間は，2つの離散的な結果であり，私たちは表や裏の確率について話をする．サイコロの標本空間は，6つの離散的な結果であり，私たちは1の目，2の目などの確率について話す．

連続的な結果空間の場合，私たちは標本空間を相互に排他的で網羅的な「区間」の有限集合に離散化することができる．例えば，人物の身長は連続スケールであるが，有限な間隔数にスケールを分割することができる．例えば1.29 m以下，1.29 mから1.35 m，1.35 mから1.40 m，1.40 mから1.45 m，．．．，2.10 m以上の各区間，というように．その後，ランダムに選択された人物が該当する区間の確率について話す．ランダムに10,000人をサンプリングし，非常に正確に身長を測定したとしよう．図4.2の上の段は10,000測定の散布図を示し，縦の破線が区間を示している．具体的には，1.60〜1.65 mの区間内に入る測定値は1,473であり，これは，その区間に入る（推定）確率は1,473/10,000 = 0.1473であることを意味している．

連続スケールのある区間に落ち込む確率のように，離散的な結果の確率は確率質量 (probability mass) と呼ばれる．大まかにいえば，「質量」という用語は，物体内の物質の量を指す．物質量が確率であり，スケールの区間がその対象である場合，質量は区間における結果の割合になる．間隔を通じて確率質量の合計は1でなければならないことに注意してほしい．

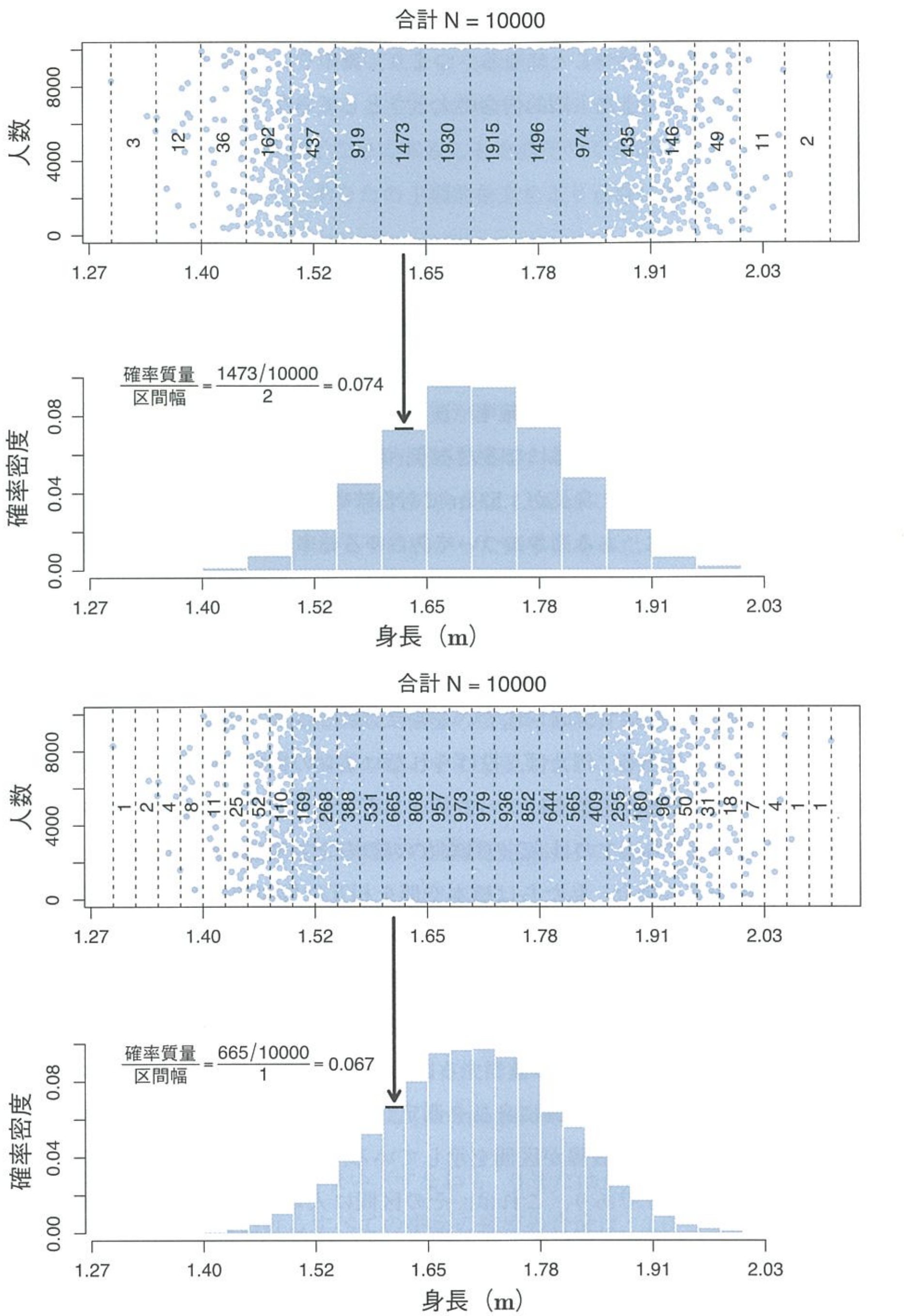

図 4.2 確率密度を計算例．2つの図それぞれにおいて，上のプロットはランダムに選択された 10,000 人の身長のばらつきを示し，下のプロットは区間 (bin) に示された特定の選択を確率密度に変換したものを示す．

4.3.2. 連続分布：密度のランデブー[5)]

連続的な結果空間について慎重に考えていくと，連続体上の間隔ではなくて，連続体のある値の確率について言及するには問題があることに気づくだろう．例えば，ランダムに

選択された人が正確に 1.707239132632... m の高さを持っている確率は，基本的に 0 である．それはどんな値についても同様である．しかし，私たちが上記の例で行ったように，間隔の確率質量について話すことはできる．間隔を使うときの問題は，しかし，その幅と端が任意であることと，広い間隔はあまり正確なものでなくなるということだ．したがってやるべきことは，間隔を無限に狭くし，各極小区間の非常に小さな確率質量について話す代わりに，間隔区間の確率質量の比率について言及することだ．この比率は確率の密度と呼ばれている．

大まかにいえば，密度は，それが占める空間の単位当たりの物質の量である．その物質の量を測定しているため，その密度は，占有量空間で割った質量である．小さな質量が高い密度を有していることに注意してほしい．すなわち，鉛 1 mg は，1 cm^3 あたり 11 g 以上の密度を有する．これは，1 mg はたった 0.000088 cm^3 の空間しか占めていないためである．重要なことは，考えるべき空間をその点の周りの微小な領域に縮小させ，その空間に対する質量の比率として，ある点の密度を考えることができるということだ．

図 4.2 は，このアイデアの例を示している．前述したように，上部は，幅 2 単位の間隔で 10,000 ランダムに選択された人々の身長（m）の散布図を示している．1.60 m から 1.65 m の間隔で平均確率密度を計算するために，区間幅によって間隔確率質量を分割する．確率質量は（推定）1,473/10,000 = 0.1473 であり，間隔の幅は 2.5 cm を 1 単位とする（すなわち，(1.65 − 1.60)/0.025 = 2.0）．したがって，区間内の平均確率密度は 0.074 である（四捨五入）．これは，間隔の間の平均確率密度である．狭い間隔の間のより正確な密度については，図 4.2 の下部を見てほしい．1.60 m から 1.625 m の間隔は，（推定）665/10,000 の質量を持っている．したがって区間の平均確率密度は (665/10,000)/((1.625 − 1.60)/0.025) = 0.067（四捨五入）である．この間隔を狭くし続けることができるし，また密度を計算し続けることができる．

図 4.2 の例では，微小区間化していない領域を通じての有限サンプルから密度を推定している．しかし，微小間隔の密度を計算するために，スケールの連続的な広がりを通じての無限母集団を考えなければならない．だから，微小空間がある 0 でない（微小空間を通じての）確率密度を含んでいたとしても，そのスケールのある点における確率密度に言及することができる．この考え方についての数学的な例をこの後に示す．

図 4.3 は別の例を示している．そこでは確率質量が 1 を超えることはできないのに，確率密度が 1 よりも大きくなっている．図 4.3 の上部は，2.13 m（213 cm）の高さになるように作られたドアからランダムに選ばれた 10,000 個について，高さをメートルで表している．普通に製造されたプロセスでも，ドアの高さはほんの少しランダムな散らばりを見せることがあるので，スケールの幅は小さいが，2.123 m から 2.143 m という幅があるという事実が図に示されている．だから，全確率質量はスケールの小さな幅の中

5)「人類史には神秘的なサイクルがある．いくつかの世代には多くの実りが与えられる．他の多くの世代は待ち望んでいる．今のアメリカ合衆国にある世代は，濃密な運命と共にある．」フランクリン・ルーズベルト，1936.

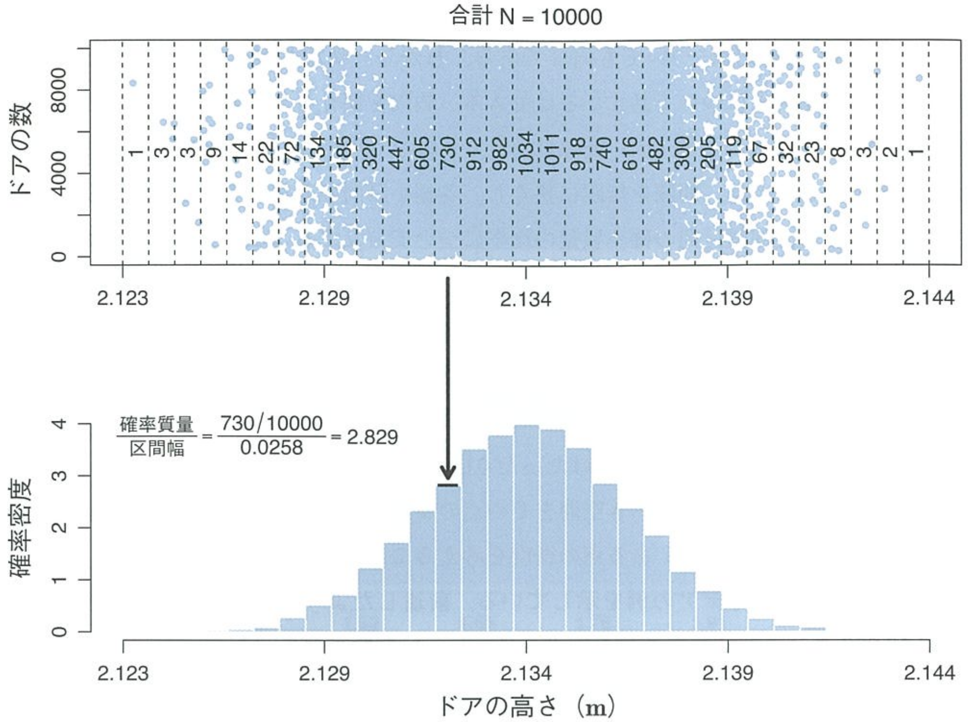

図 4.3　1.0 より大きい確率密度の例．ここでは，すべての確率質量は，スケールの小さな領域に集中し，また密度は，スケールのいくつかの値で高くすることができる．密度の注釈付きの計算は，表示のため丸めた区間制限を使用する．

に集中している．結果 2.143 m 付近の確率密度は 1.0 を超える．例えば，2.13130638 から 2.1319617 の間隔では，730/10,000 = 0.073 の確率質量がある．しかし，この質量は，たったの (2.1319617 − 2.13130638)/0.025 ≒ 0.0258 の区間幅の上に集中している．したがって，区間内の平均密度は 0.073/0.0258 = 2.829 である．1.0 より大きい確率密度については不思議なことは何もない．それはスケールに対して確率質量が相対的に高濃度であることを意味するだけである．

4.3.2.1.　確率密度関数の特性

一般的に，任意の連続値も間隔に振り分けられるため，間隔の確率質量の総和は 1 でなければならない．なぜなら，測定の定義によって測定スケールの特定の値が生じるからだ．その事実を式に書くことができるが，まずのいくつかの表記を定義する必要がある．連続変数は x と表記する．x 上の区間の幅は，Δx（"Δ" というはギリシャ文字，「デルタ」δ の大文字である）と表記する．i を区間の添え字とすると，$[x_i, x_i+\Delta x]$ は x_i から $x_i+\Delta x$ の区間を表す．i 番目の区間の確率質量は $p([x_i, x_i+\Delta x])$ と表記する．以下のように，それらの確率質量の合計が 1 でなければならない．

$$\sum_i p([x_i, x_i+\Delta x]) = 1 \tag{4.1}$$

ここで確率密度の定義を思い出そう．すなわち，それは区間の幅の確率質量の比である．各区間の密度の観点から，Δx で分割して乗算することにより，式 4.1 を次のように書き換えることができる．

$$\sum_i \Delta x \frac{p([x_i, x_i+\Delta x])}{\Delta x} = 1 \tag{4.2}$$

区間幅が無限小になるように制限を課すと，x の周りの間隔の幅を Δx の代わりに dx と表すことができ，x まわりの微小間隔における確率密度を，単に $p(x)$ と表す．確率密度 $p(x)$ を，$p([x_i, x_i+\Delta x])$ と混同しないように．それはその間隔における確率質量である．そして，式 4.2 での総和は積分となる．

$$\underbrace{\sum_i}_{\int} \underbrace{\Delta x}_{dx} \underbrace{\frac{p([x_i, x_i+\Delta x])}{\Delta x}}_{p(x)} = 1 \quad \text{つまり，} \int dx\, p(x) = 1 \tag{4.3}$$

本書では，式 4.3 のように，積分は式の右端におく代わりに積分記号の次の項に dx と書く．この配置は，従来の表記ではないが，それは間違っているわけでも，本書独自の記法でもない．積分記号の dx をこのように配置するのは，どの変数が積分されていくかを見やすくするものであり，積分記号に添え字を置かなくてもよいからである．複数の変数を伴う関数の積分が生じた場合には，この使用法は特に役に立つ．積分記号の隣に dx をおくことで，離散和と積分を書きなおすときも，グルーピングする項を維持するので，$\sum_x$ は $\int dx$ として dx を式の最後に動かすことなく表現できる．

繰り返しになるが，式 4.3 における $p(x)$ は，x の周囲の微分区間における確率密度である．普通，確率質量か確率密度かを参照しなくても文脈でわかるようになるし，どちらのときも同じ表現である $p(x)$ で示す．例えば，x がサイコロの 6 面の値のとき，$p(x)$ は確率質量である．あるいは x が身長の正確な点の値である場合，$p(x)$ は確率密度である．しかし，使用上の「ずれ」がある場合もある．例えば，x が身長であるが，そのスケールが区間に分割されている場合，$p(x)$ は本当はその x が落ち込む区間の確率質量を参照していることになる．最終的には，文脈に注意し，多少の曖昧さを許容しなければならないだろう．

4.3.2.2.　正規分布の確率密度関数

非負の値だけからなり，積分して 1 になる関数（式 4.3 もそれを満たしている）は，確率密度関数と解釈することができる．おそらくもっとも有名な確率密度関数は，ガウス分布としても知られている正規分布である．正規曲線のグラフは，よく知られている釣鐘型である．例を図 4.4 に示す．

正規分布の確率密度を表す数式には，2 つのパラメータがある．μ（ギリシャ文字の「ミュー」）は分布の平均を表し，σ（ギリシャ文字の「シグマ」）は，分布の標準偏差 (standard deviation; SD) を示す．μ の値は，釣鐘の中央の x 軸上に位置しているので，

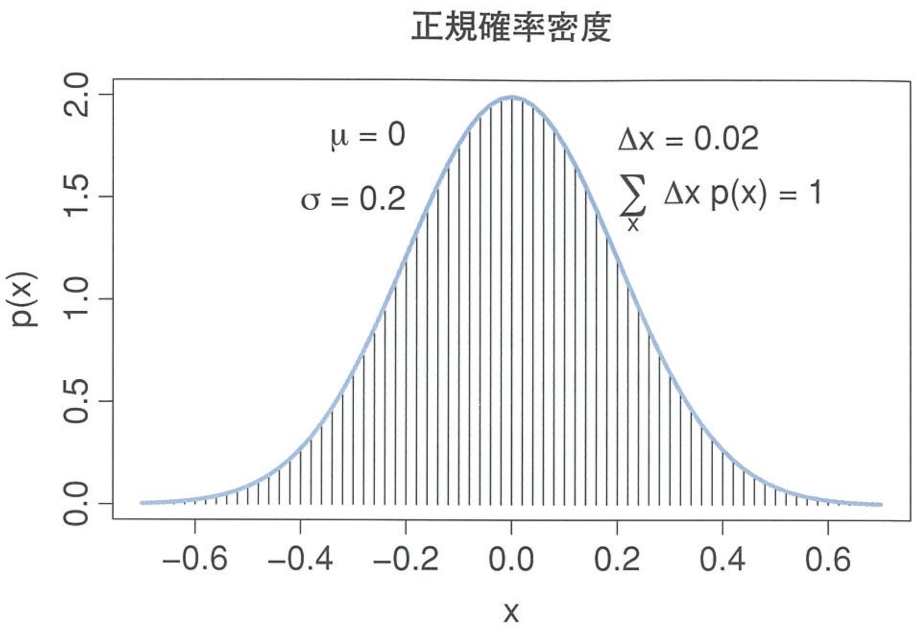

図 4.4 狭い間隔のくしで示す正規分布の確率密度関数．積分は，各区間の幅と高さの積を合計することによって近似される．

位置パラメータと呼ばれ，σ の値は，ベルの幅がどれくらいかを示すので，スケールパラメータと呼ばれる．セクション 2.2 で説明したように，パラメータについて分布の位置とスケールを調整することのできるつまみのようなものと考えることができる．正規分布の確率密度の数式は，次の通りである．

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) \tag{4.4}$$

図 4.4 は正規分布における μ と σ の特定の値の一例を示している．標準偏差 σ が小さいとき，確率密度のピークが 1.0 よりも大きくなることに注意しよう．つまり，標準偏差が小さい場合，確率質量の多くが狭い区間に押し込められ，その結果，その区間の確率密度が高くなる．

図 4.4 は，正規分布の曲線の下の面積は，実際，1 であることを示している．x 軸は，狭い幅を持った密な区間 Δx に分割される．正規分布の密度の積分は，式 4.2 のようにすべての小さな間隔の質量を合計することによって近似される．グラフ内のテキストに見られるように，間隔の面積の和は，実質的に 1.0 である．丸め誤差と，分布の両端が合計に含まれてないことだけが，ちょうど 1 にならない理由である．

4.3.3. 分布の平均と分散

確率 $p(x)$ で生成された（カテゴリカルではない）数値 x があるとき，繰り返し x の値をサンプリングしたら，長期的にはその平均値がどのようになるか疑問に思うだろう．例えば，公平な 6 面あるサイコロを持っている場合，その 6 つの値のそれぞれが，長期

的に 1/6 ずつ出るはずだろう．そのためサイコロの長期的な平均値は $(1/6)1 + (1/6)2 + (1/6)3 + (1/6)4 + (1/6)5 + (1/6)6 = 3.5$ となる．別の例として，勝つと 100 ドルもらえ，その確率が 0.001 であるスロットマシンをプレイする場合，0.14 の確率で勝ち 5 ドルをもらえ，負ければ 1 ドル失うとすれば，長期的な私たちの利得は (0.001)(100 ドル) + (0.14)(5 ドル) + (0.859)(−1 ドル) = −0.059 ドル となる．言い換えれば，長い目で見ると約 6 セントずつ失っていくのである．これらの計算は何をしたのか，考えてみよう．私たちは，それが起こる確率によってそれぞれが起こりうる結果を重みづけした．この手順は，期待値と呼ばれる確率分布の平均を定義することであり，$E[x]$ で示される．

$$E[x] = \sum_x p(x)\, x \tag{4.5}$$

x の値が離散的である場合，式 4.5 が適用され，そのため，$p(x)$ は確率質量を表す．x が連続量であるとき，$p(x)$ は確率密度を示し，合計が無限小区間で積分される．

$$E[x] = \int \mathrm{d}x\, p(x)\, x \tag{4.6}$$

概念的には，x が離散的であっても連続的であっても同じであり，$E[x]$ は値の長期的な平均値である．

直観的にいうと，分布の平均値は，一般的に分布の中央付近に位置している．例えば，正規分布の平均値は，そのパラメータ μ の値であることがわかる．つまり，この場合，$E[x]$ が μ であることが判明した．このことを具体的な例で示したのが図 4.4 で，分布の大部分は，$x = \mu$ 上でセンタリングされていることがわかる．μ の正確な値については，図中のテキストを参照してほしい．

ここでは式 4.6 を使用して，連続分布における平均値の計算の例を示す．間隔 x が $[0,1]$ 上で定義された確率密度関数 $p(x) = 6x(1-x)$ を考えてみよう．これはまぎれもなく確率密度関数である．なぜなら，その逆さまになった放物線は，$x = 0$ から始まり，$x = 0.5$ でピークに達し，そして $x = 1$ で再びベースラインに落ちる．それは対称的な分布であるため，直観的に平均は，$x = 0.5$ であり，平均点が中間の点であることを教えてくれる．実際に確認してみよう．

$$\begin{aligned}
E[x] &= \int \mathrm{d}x\, p(x)\, x \\
&= \int_0^1 \mathrm{d}x\; 6x(1-x)\, x \\
&= 6\int_0^1 \mathrm{d}x\; \left(x^2 - x^3\right) \\
&= 6\left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_0^1 \\
&= 6\left[\left(\frac{1}{3}1^3 - \frac{1}{4}1^4\right) - \left(\frac{1}{3}0^3 - \frac{1}{4}0^4\right)\right] \\
&= 0.5
\end{aligned} \tag{4.7}$$

本書ではほとんど計算しないし，式 4.7 はこれから出るものの中でも高度な計算である．もしあなたの計算についての知識が錆び付いていたとしても，概念的に理解するために読み続けてほしい．

確率分布の分散は，平均値から離れている数値を表す．x の値が平均値からどれくらい離れているかについては色々な定義がありうるが，特に「分散」という言葉を使うときは，x と平均値との差の二乗に基づいている．分散の定義は，平均からの x の値の偏差平方の平均 (mean squared deviation; MSD) である．

$$\mathrm{var}_x = \int \mathrm{d}x\, p(x)\ (x - E[x])^2 \tag{4.8}$$

式 4.8 は平均の式 (式 4.6) とよく似ているが，x の代わりに x の確率で重みづけられたところだけが違っており，x の確率によって重みづけられた $(x - E[x])^2$ の積分となっている．言い換えれば，分散は $(x - E[x])^2$ の平均値そのものである．離散分布の場合は，式 4.8 の積分は，式 4.5 と式 4.6 との間の関係に類似した総和となる．分散の平方根は，平均平方偏差 (root mean squared deviation; RMSD) を意味しており，分布の**標準偏差**と呼ばれている．

正規分布の分散は，そのパラメータ σ の値の平方であることが判明した．このように，正規分布では $\mathrm{var}_x = \sigma^2$ である．言い換えると，正規分布の標準偏差は，パラメータ σ の値である．正規分布では，分布の約 34% が μ と $\mu + \sigma$ の間にある（エクササイズ 4.5 を参照）．図 4.4 を見て，μ と $\mu + \sigma$ は x 軸上のどこにあるか確認し（μ と σ は図中に文字として示されている），1SD（標準偏差ひとつ分）が平均からどれくらいのところにあるかについて視覚的な印象を持っておこう．ただし，他の形状の分布に一般化しすぎないようにしよう．すなわち，非正規な分布は，それらの平均と 1 つ目の標準偏差との間に非常に異なる領域を持つことがありうる．

確率分布は，測定値またはパラメータ値の確率を参照することができる．確率は生成的なプロセスからある値がどれほど多くサンプリングされたか，または値が他の値と比較してどのくらいの確信度であるかを意味するものと解釈することができる．$p(\theta)$ が θ の確信度を表す場合，サンプリングされた θ の確率の代わりに，$p(\theta)$ を θ の典型的な確信度を表していると考えることもできる．θ の標準偏差は，その分布の広さを測っているのだが，候補となる値を通じての不確かさの測度だと考えることもできる．標準偏差が小さい場合，θ は平均周りの値であると強く信じられる．標準偏差が大きい場合，θ の値がどれくらいであるかはっきり信じられなくなる．標準偏差が不確かさを表しているというこの考えは，この後，何度も繰り返し出てくる．分布の幅に関する他の測度は，次に説明するように，最高密度の間隔である．

4.3.3.1. 分散を最小化する平均

もう 1 つの考え方を今から紹介するが，平均と，それが分散の定義にどう関係するかというところから始める代わりに，分散の定義と平均の定義から得られるものの話から

始めよう．この代替概念のもとでは，確率分布の**中心傾向**の値を定義することが目標である．ある値が分布の中心傾向を表しているというのは，分布において高い確率を持つ値に近いときである．つまり，分布の中心傾向を，それと他のすべての x の値との間の，長期にわたって期待される距離を最小化する何らかの M と定義する．しかし，どのように，値の間に「距離」を定義できるだろうか？　距離を定義する 1 つの方法は差の平方である．すなわち，x と M の距離は $(x - M)^2$ である．この定義の長所は，M から x にかけての距離と x から M にかけての距離が同じであるということである．つまり，$(x - M)^2 = (M - x)^2$ である．しかし，この定義の一番の長所はこれ以降の代数の多くが扱いやすくなるということである（それはここでは繰り返さないが）．それゆえ，中心傾向は $(x - M)^2$ の期待値を最小化する値 M である．したがって，$\int \mathrm{d}x\, p(x)\, (x - M)^2$ を最小にする M が求める答えである．見覚えがあるだろうか．これは，式 4.8 で，基本的に分布の分散のための式だが，ここでは M の関数として考える．結果は次の通り．$\int \mathrm{d}x\, p(x)\, (x - M)^2$ を最小にする M の値は $E[x]$ である．言い換えると，分布の平均値は，予想される平方偏差を最小にする値である．つまり，平均値は分布の中心傾向である．

余談として，M と x との間の距離が $|x - M|$ として定義されている場合，期待される距離を最小化する値を分布の**中央値**と呼ぶ．同様に，すべてがぴったり一致していれば 0，いずれかが不一致であれば 1 と定義された距離を用いて，分布の表現に用いるのが**最頻値**である．

4.3.4. 最高密度区間 (HDI)

分布を要約する別の方法でよく使われるものが，最高密度区間，略して HDI(the Highest Density Interval) である[6]．HDI は分布のもっとも信頼できる部分，かつ，分布の大部分をカバーしているところを示す．このように，HDI は分布の大部分をまたがる区間，例えば 95% を指定して分布をまとめたもので，区間内部のすべての点が区間外の任意の点よりも高い信頼性を有している．

図 4.5 は，HDI の例を示す．上部は，平均 0 と標準偏差 1 の正規分布を示している．この正規分布は 0 を中心に対称となっているため，95%HDI は −1.96 から 1.96 にまで及ぶ．図 4.5 の灰色で塗られているこの曲線の区間内領域は，0.95 の領域である．また，これらの区間内の任意の x の確率密度は，これらの区間外の x より高い確率密度を有す．

図 4.5 の中央部は，歪んだ分布の 95%HDI を示す．定義上，95%HDI 区間の曲線下面

---

6)　本によっては，最高密度領域を表す言葉である HDR(the Highest Density Region) として HDI を示しているものがある．領域，という言葉は多次元を表現できるが，区間のほうは 1 次元しか表現できない．本書ではほとんどいつも 1 時点の 1 つのパラメータの HDI についての話なので，私は混乱をなくすために "HDI" という言葉を使おうと思う．また，本によっては，最高確率密度を表す HPD(Highest Probability Density) として HDI に言及することもあるが，私がそれを使わないのは，"HDI" よりも "HPD 区間" と書くのに場所をとられるからだ．さらに本によっては，最高事後確率密度を表す HPD(Highest Posterior Density) という言葉を HDI として使う人もいるが，私がそれを好んで使わないのは事前分布も HDI を持っているからだ（訳注；Highest Prior Density と区別がつかない）．

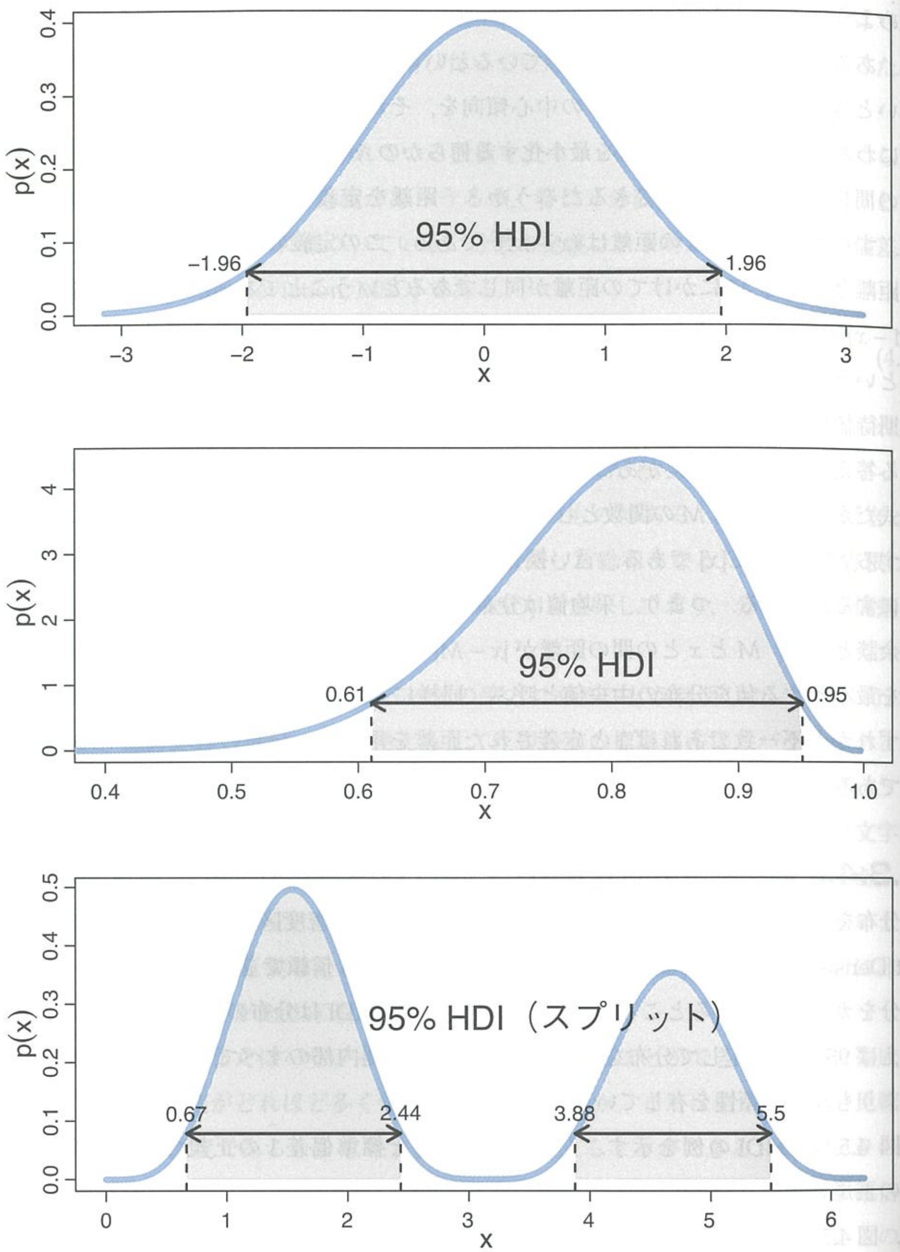

図 4.5　95% 最高密度区間 (HDI) の例．各例において，区間内のすべての x 値は，区間外の x 値よりも高い密度を有し，かつ間隔の内部点の合計質量が 95% である．95% の領域は影をつけ，それは水平方向の矢印下のゾーンを含む．水平方向の矢印は，95%HDI の幅を示しており，その両端部に x の値が示されている（四捨五入している）．水平方向の矢印の高さは，95%HDI 内のすべての x の値が，少なくともその密度を超えることを示している．

積（図の灰色で示した箇所）は 0.95 の領域を有している．そして，これらの区間内の任意の x の確率密度は，これらの区間外の x よりも高い．重要なのは，左 HDI 限界未満である左側の分布の裾は，右 HDI の限界値を超えている右側の分布の裾より大きくなっている点であることに注意してほしい．言い換えれば，HDI は必ずしも HDI 領域の外を均

等にするものというわけではない（以前に両裾が等しい確信区間に遭遇した人は，HDI が両裾が等しい区間とはどう異なるかを理解するために，p. 348 の図 12.2 を先に見てもよい）．

図 4.5 の下部は，非現実的な二峰性の確率密度関数を示す．多くの現実的な適用では，このような多峰分布は発生しないが，この例では，HDI の定義を明確にするために有用である．この場合，HDI は 2 部分区間，分布の各最頻値について分割される．しかし，定義された特性は前と同じである．図中の灰色で塗られた 95%HDI の範囲内で曲線下の領域，0.95 の総面積を有しており，この範囲内の任意の x は外側の任意の x より高い確率密度を有する．

HDI の正式な定義は，2 つの本質的な特性の数学的表現である．95% HDI は少なくともある値 W よりも大きな密度を持つであろう x のすべての値を含んでおり，こうしたすべての x について積分した値が 95% なのである．数式的には，95% HDI の x の値は，$p(x) > W$，ここで W は $\int_{x:p(x)>W} \mathrm{d}x\, p(x) = 0.95$ を満たす．

分布が値の確信度を意味しているのであれば，HDI の幅は信念の不確実性を測定する別の方法である．HDI が広い場合には，信念が不確実である．HDI が狭い場合には，信念が比較的一定である．第 13 章で詳細に説明するように，研究の目的は，特定のパラメータ値に関する確実性を，かなり高い度合いで達成するためのデータを得ることである．望ましい確実度は，95%HDI の幅として測定することができる．もし μ が，薬によってどのくらい血圧を下げるかを測るものとするなら，研究者は，血圧計の長くとも 5 目盛り内に 95%HDI の幅で推定したいだろう．別の例として，θ が候補者 B より候補者 A を選好する人口の測度だとすると，研究者は少なくとも 10 ポイント内に 95% HDI を推定したいと思うだろう．

4.4.　2 次元分布

私たちは 2 つの結果の連動について関心があるという状況が多々ある．クイーンとハート，両方のカードが配られる確率はどれくらいか？　赤い髪で緑の目の人に会う確率はどれくらいか？　サイコロとルーレットを含むボードゲームをプレイするとき，サイコロとルーレットの両方が公平であることについて，私たちはある程度の確信度を持っているのではないか．

これらのアイデアを発展させるための具体的な例として，表 4.1 に人の目の色と髪の色の様々な組み合わせの確率を示し，これについて考えてみよう．このデータは便利なサンプル事例から取ってきたもので (Snee, 1974)，任意の大きな集団の代表であることを意味するものではない．表 4.1 は，行に 4 種類の考えられる目の色が，列に 4 種類の考えられる髪の色を示す．表中の各セルの数字は，目の色と髪の色の特定の組み合わせの同時確率を示している．例えば，左上のセルは，茶色の目と黒髪の同時確率は 0.11（すな

表 4.1　髪と目の色の組み合わせ比率

	髪の色				
目の色	黒	茶色	赤	ブロンド	周辺確率（目の色）
茶色	0.11	0.20	0.04	0.01	0.37
青	0.03	0.14	0.03	0.16	0.36
ハシバミ色	0.03	0.09	0.02	0.02	0.16
緑	0.01	0.05	0.02	0.03	0.11
周辺確率（髪の色）	0.18	0.48	0.12	0.21	1.0

行や列の中にはその周辺と合計が一致しないものがあるが，これはもとのデータの丸め誤差によるものである．データは Snee (1974) より．

わち，11%）であることを示す．目の色と髪の色の組み合わせはすべて等しく起こりうるというわけでないことに注意してほしい．例えば，青い目と黒い髪の同時確率はたったの 0.03（すなわち，3%）である．目の色 e，髪の色 h の同時確率を $p(e,h)$ として示す．同時確率の表記法は対称である．すなわち $p(e,h) = p(h,e)$ である．

目の色全体についての確率に興味があるのであれば，髪の色についてはまとめてしまいたいと思うかもしれない．そうした確率は表の右側に示されているが，それらは，周辺確率と呼ばれる．行和の計算は，それぞれの行の同時確率を合計するだけでよい．例えば，緑の目の周辺確率は，髪の色と関係なく 0.11 である．表に含まれる同時確率の値は，もとのデータから丸め誤差が生じているので，総和しても表示された周辺度数とぴったり一致しない．目の色 e の周辺確率は $p(e)$ と表記され，それは髪の色の全体の同時確率を合計することにより計算される．すなわち，$p(e) = \sum_h p(e,h)$ である．

もちろん，同様に様々な髪の色の周辺確率を考慮することができる．髪の色の周辺確率は，表 4.1 の下の余白に表示されている．例えば，黒い髪の確率は，目の色と関係なく 0.18 である．周辺確率は，各列内の同時確率を合計することによって計算される．したがって，$p(表) = \sum_e p(e,h)$ である．

一般的には，行変数 r と列変数 c を考える．行変数が離散ではなく連続的なとき，$p(r,c)$ は確率密度である．そして，周辺確率を計算するための総和は，積分 $p(r) = \int \mathrm{d}c\, p(r,c)$ となる．そこで得られた周辺分布 $p(r)$ もまた，確率密度である．この総和していく過程は，c の周辺化や変数 c の統合と呼ばれる．もちろん r の周辺化によって確率 $p(c)$ を決定することができる．すなわち，$p(c) = \int \mathrm{d}r\, p(r,c)$ である．

4.4.1．条件付き確率

ある1つの結果が真であるということを知っているとすると，私たちはその上でのもう1つの結果の確率を知りたい，と考えることもあるだろう．例えば，表 4.1 の母集団からランダムに人をサンプリングする．ここで私があなたに，この人は青い目であると伝えたとしよう．その情報に関する条件下で，その人がブロンドの髪である確率（または任意の他の特定の髪の色）はどれくらいだろうか？　答えを計算する方法は直観的で明確で

表 4.2　条件付き確率の例

	髪の色				
目の色	黒	茶色	赤	ブロンド	周辺確率（目の色）
青	0.03/0.36 =0.08	0.14/0.36 =0.39	0.03/0.36 =0.08	0.16/0.36 =0.45	0.36/0.36=1.0

表 4.1 における青い目の人々の中で，各髪の色 h の比率はどれくらいか．各セルは $p(h|青) = p(青,h)/p(青)$ を少数第2位までを示している．

ある．私たちは表 4.1 の青い目の行からその合計（すなわち，周辺確率）の青い目の人は 0.36 であり，青い目とブロンドの髪の人数は 0.16 である．したがって，0.36 の青い目のうち，0.16/0.36 がブロンドの髪をしている．言い換えれば，青い目の人の 45% はブロンドの髪である．また，青い目の人の，0.03/0.36 = 8% は黒い髪をであることに注意してほしい．表 4.2 は，それぞれの髪の色について，この計算を示している．

髪の色の確率は，それぞれ起こりうる髪の色の確信度を表す．表 4.1 の周辺分布からわかるように，この人々の群では，ブロンドの髪を持つ一般的な確率は，0.21 である．しかし，ある人が青い目の群から来たことがわかっているとき，この人がブロンドの髪をしている確信度は，表 4.2 から見てとれるように，0.45 に上昇する．この髪の色の確率についての確信度が更新されることこそ，ベイズ推論なのである！　しかしここではさらに話を進めよう．次の章ではベイズ推論の基本的数学についてより詳しく解説する．

条件付き確率のための直観的な計算は，簡単な公式で表すことができる．目の色のときの髪の色の条件付き確率を $p(h|e)$ として示す．これは，「e が与えられたときの h の確率」という．上記の計算は，$p(h|e) = p(e,h)/p(e)$ とも書かれる．この式を，条件付き確率の定義とする．周辺確率は，単にセルの確率の総和であることを思い出すと，定義は $p(h|e) = p(e,h)/p(e) = p(e,h)/\sum_h p(e,h)$ と書くことができる．分子の h は髪の色の特定の値であるが，分母の h は，髪の色のすべての起こりうる値をとる変数であるので，その式は混乱するかもしれない．h の2つの意味を明確にするために，$p(h|e) = p(e,h)/p(e) = p(e,h)/\sum_{h^*} p(e,h^*)$ と書きなおすこともできる．ここで h^* は髪の毛の色のありうる値を意味している．

条件付き確率の定義は，より一般的な記号として，行を参照する r と列を参照する c を使って書きなおすことができる．さらに，離散的な値を持つ属性については，条件付き確率は以下のように定義される．

$$p(c|r) = \frac{p(r,c)}{\sum_{c^*} p(r,c^*)} = \frac{p(r,c)}{p(r)} \tag{4.9}$$

列属性が連続である場合，和は積分となる．

$$p(c|r) = \frac{p(r,c)}{\int \mathrm{d}c\, p(r,c)} = \frac{p(r,c)}{p(r)} \tag{4.10}$$

もちろん他の変数で条件づけることもできる．すなわち，$p(r|c)$ を $p(c|r)$ の代わりに考えることもできる．一般に，$p(r|c)$ は $p(c|r)$ と同じではないことを理解しておくことが大事である．例えば，雨が降っていることによって地面が濡れている確率は，地面が濡れているときに雨が降っている確率とは異なっている．次の章では，$p(r|c)$ と $p(c|r)$ との関係の拡張の議論をする．

条件付き確率には時間的順序が存在しないことを認識することも重要である．「y が与えられたときの x の確率」は，y がすでに起こっている，または x はまだ起こっていないということを意味するものではない．起こりうる結果の特定のサブセットの確率の計算に限定していることを表しているだけだ．$p(x|y)$ についてのよりよい説明としては，「値 y と同時に起こるあらゆる結果において，x という値でもある割合」である．だから，例えば，朝に雲があるとき前日の夜に雨が降った条件付き確率について話すことができる．これは単に，前の晩に雨が降ったすべての曇った朝の割合について語っているのである．

4.4.2. 属性の独立

6面体ダイス（サイコロ）とコインがあるとする．それらは公平であると仮定する．コインを投げると，表が出た．コインのこの結果から考えると，ダイスで3の数字が出る確率はどれくらいだろう？　この質問に答えるときは，おそらく「コインがサイコロに影響を与えることはないので，コインの結果に関係なく，3が出るサイコロの確率は1/6である」と思うだろう．あなたがそう考えたのなら，コインとサイコロは独立していると想定していた，ということだ．

一般的に，y の値が x の値に影響を及ぼさない場合，y が与えられたときの x の確率は一般的な x の確率にすぎない．式で記述するなら，この考えは x と y のすべての値に対して，$p(x|y) = p(x)$ と表される．それが意味するものについて考えてみよう．式4.9または式4.10の条件付き確率の定義から $p(x|y) = p(x, y)/p(y)$ ということがわかっている．式を組み合わせることにより，x と y のすべての値に対して，その $p(x) = p(x, y)/p(y)$ を意味する．$p(y)$ を両辺に掛けると，x と y のすべての値について $p(x, y) = p(x)p(y)$ になるということがわかる．逆もまた真なり，である．すなわち x と y のすべての値において $p(x, y) = p(x)p(y)$ のとき，x と y のすべての値に対して $p(x|y) = p(x)$ となる．したがって，これらの条件のいずれかは，属性の独立性の数学的な定義である．繰り返しになるが，属性 x と y は，独立している．つまり，x の y すべての値について，$p(x|y) = p(x)$ は $p(x, y) = p(x)p(y)$ と数学的に等価である．

目の色と髪の色について表4.1 (p.94) の例を考えてみよう．属性は独立しているだろうか？　日常の経験から直観的に，私たちは，答えがNoであることを知っているが，私たちは数学的にそれを表現することができる．独立性を反証するために必要なのは，ある目の色 e とある髪の色 h が $p(h|e) \neq p(h)$ である，または同じことだが $p(e, h) \neq p(e)p(h)$，ということである．私たちは青い目とブロンドヘアのように，すでにこのようなケースを扱ってきた．表4.1は，青い目でかつブロンドの周辺確率が $p(\text{ブロンド}) = 0.21$ であることを示している．表4.2は，ブロンドの条件付き確率は $p(\text{ブロンド} \mid \text{青}) = 0.45$ であることを示している．したがって，$p(\text{ブロンド} \mid \text{青}) \neq p(\text{ブロンド})$ となる．私たちは，周辺確率をたすき掛けして，それと同時確率とが等しくないことを示すことによって，独立性を反証することができる．具体的には，$p(\text{青}) \cdot p(\text{ブロンド}) = 0.36 \cdot 0.21 = 0.08 \neq 0.16 = p(\text{青, ブロンド})$ となる．

独立した2つの属性の簡単な例として，標準的なデッキでスート[c]とトランプの数字を考えてみよう．4つのスート（ダイヤモンド，ハート，クラブ，スペード）があり，かつ，それぞれのスートには13の値（エース，2，．．．，9，ジャック，クイーン，キング）があって，52枚のカードを構成している．ランダムにカードが配られたとしよう．それがハートである確率はいくつだろう？　（答え：13/52 = 1/4）　あなたに見せないように私がそのカードを見るとクイーンであり，私がそれはクイーンだったとあなたに伝えたとする．今，それがハートである確率はいくつだろう？　（答え：1/4）．一般的に，あなたのカードの数字を伝えることはスートの確率を変更しないので，値およびスートは独立している．周辺確率を掛け合わせることでこれを確認することができる．すなわち，数字とスートのそれぞれの組み合わせは，1/52の確率でその手札にある（公平にデッキがシャッフルされていれば，である）．1/52は正確には，いずれかの数字 (1/13) のいずれかのスート (1/4) 倍の周辺確率であることに注意してほしい．

他の文脈では，複数のパラメータについての信念の数学的記述を構築しているときに独立性が出てくるだろう．私たちはあるパラメータへの信念について数学的な記述を作成し，そして他のパラメータへの信念については別の数学的な記述を作成するだろう．その後，パラメータの組み合わせに関する信念を記述するために，独立性を仮定することがよくある．そして同時確信度を特定するためには，個別の確信度を掛けあわせるだけでよい．

4.5. 付録：図4.1のためのRコード

図4.1は，スクリプト `RunningProportion.R` によって製作された．それを実行するには，（ワーキングディレクトリにファイルが含まれていることを仮定して）Rのコマンドラインで `source("RunningProportion.R")` を入力する．擬似乱数生成器なので，実行するたびに得られる結果は異なる．特定の開始状態に擬似乱数生成器を設定したい場合は，`set.seed` コマンドを使用する．図4.1の例では，その後 `set.seed(47405)` と `source("RunningProportion.R")` を入力する．

あなたは，グラフのために作成されたウィンドウのサイズを制御し，その後，それを保

c)　訳注：トランプのマークのこと．