

Contents

1	条件付き確率の一般論	2
1.1	簡単な復習	2
1.2	例1：古典的な場合の条件付き確率	4
1.3	例2：一般の場合の条件付き確率	5
1.4	ベイズの定理	6
2	決定木	7
2.1	動機づけの例：ロッククライマー	7
2.2	例：ジャングルと泥棒	9
2.3	ロッククライマーに戻る	11
2.4	演習：復元なしの抽出	12
3	有名な例	14
3.1	モンティ・ホール問題	14
3.2	誕生日のパラドックス	15
3.3	医療検査	17

この講義では、条件付き確率を発展させる。第5回講義で簡単に見たように、条件付き確率とは、扱っているシステムについて何らかの追加情報を得た後に計算される確率である。これから、古典的確率の外にある一般的な計算をいくつか見ることで、この概念をより丁寧に学ぶ。さらに、多段階の確率問題を整理するための実用的な方法である 決定木 も導入する。最後に、条件付き確率に関わる三つの有名な例、モンティ・ホール問題、誕生日のパラドックス、そして医療検査について議論する。

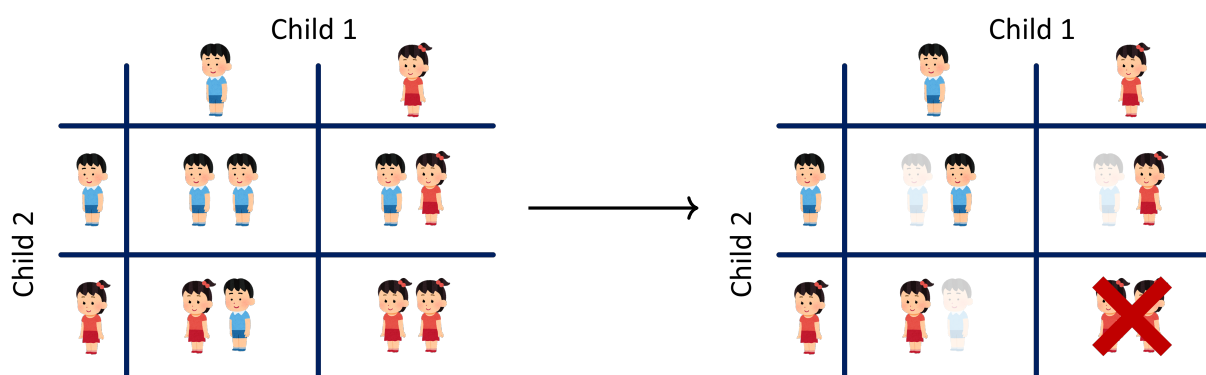
1 条件付き確率の一般論

あなたの隣人には子どもが二人いるとする。あなたはその家に行き、ドアをノックすると、男の子がドアを開けた。このとき、もう一人の子が男の子である確率はいくらだろうか。

驚くべきことに、答えは実は $\frac{1}{3}$ である。初めて聞くとかなり意外に感じられるが、これは確かに正しい。問題を簡単に分解してみよう。まず、可能な順序付きの結果は四つある：

$$S = \{BB, BG, GB, GG\}.$$

素朴には、これらの結果はそれぞれ $\frac{1}{4}$ の確率で起こると考えられる。しかし、少なくとも一人の子が男の子であることを見たことで、情報が更新される。つまり、「GG」という選択肢はもはや不可能であることが分かる。したがって、可能性の空間は次のように減少する：



言い換えると、標本空間は実質的に次の集合へと減少する：

$$\{(\text{男の子}, \text{男の子}), (\text{女の子}, \text{男の子}), (\text{男の子}, \text{女の子})\}.$$

この三つの結果のうち、二人とも男の子であるものは一つだけである。したがって、

$$P(\text{二人とも男の子} \mid \text{少なくとも一人が男の子}) = \frac{1}{3}.$$

ここで、条件付き確率の計算を一つ実演したことになる。条件付き確率とは、情報の更新や追加の仮定に基づいて不確実性を計算するものである。このように、条件付き確率の結果はしばしば驚くほど直感に反する。今日は、この混乱しやすい状況を扱うためのいくつかの方法を導入する。まず、これまで数回の講義で見た理論を簡単に復習し、それを基本的な計算へ応用する。その後、確率を非常に分かりやすい結果の体系へ整理できる決定木を扱う。最後に、条件付き確率に関する直感に反する有名な例、モンティ・ホール問題、誕生日のパラドックス、医療検査の妥当性で締めくくる。

1.1 簡単な復習

古典的確率では、すべての個別の結果は同様に起こりやすい、ということを思い出そう。第6回講義で議論したように、これは確率を決めるための特別な仮定であり、次のような特有の方法を与える：

$$P(A) = \frac{|A|}{|S|}.$$

上の公式は、すべての結果が同様に起こりやすいと仮定されているため、任意の事象 A の大きさを単に数え上げ、標本空間 S の中で A が占める割合を計算することで $P(A)$ を求められる、ということ述べている。

対照的に、一般の確率を扱うときには、すべての結果が同じ確率で起こるという仮定を緩める。個別の結果を ω と書くと、一般の確率では、すべての確率 $p(\omega)$ が同じであるとは仮定しない。その結果、事象 A の大きさを単に数え上げるだけでは、その確率を求めることはできない。代わりに、 A に含まれる結果の確率を足し合わせなければならない。この和が、 $P(A)$ という確率になる：

$$P(A) = \sum_{\omega \in A} p(\omega).$$

前回の講義で述べたように、複合事象に関する公式は、一般の場合でも有効である。ただし、その正当化は異なる。それでも、期待通りに用いることができる。

1. 事象 A が起こらない確率を求めるには、次の公式を用いればよい：

$$P(A \text{ でない}) = 1 - P(A).$$

2. 二つの事象 A と B が同時に起こる確率を求めるには、次のいずれかの公式を用いればよい：

$$P(A \text{ かつ } B) = \begin{cases} P(A) \cdot P(B), & A \text{ と } B \text{ が独立である場合,} \\ P(A|B) \cdot P(B), & A \text{ と } B \text{ が従属であり、ただし } P(B) \neq 0 \text{ の場合.} \end{cases}$$

3. 二つの事象 A または B のどちらかが起こる、あるいは両方起こる確率を求めるには、次のいずれかの公式を用いればよい：

$$P(A \text{ または } B) = \begin{cases} P(A) + P(B), & A \text{ と } B \text{ が互いに排反である場合,} \\ P(A) + P(B) - P(A \text{ かつ } B), & A \text{ と } B \text{ が互いに排反でない場合.} \end{cases}$$

第5回講義で、記号 $P(A|B)$ は、 B が確実に起こると分かっているときに A が起こる確率を表す、と述べた。その議論では、古典的確率の場合、 $P(A \text{ かつ } B)$ の公式から $P(A|B)$ の公式を発見できることを見た。古典的な Golden Formula に、意図的に $1 = \frac{|B|}{|B|}$ を挿入すると、次のようになる：

$$P(A \text{ かつ } B) = \frac{|A \cap B|}{|S|} = \frac{|A \cap B|}{|S|} \cdot \frac{|B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{|B|}{|S|}.$$

第二の因子はちょうど $P(B)$ であり、第一の因子が新しい公式を与える：

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

確率論的に言えば、この公式は、 S ではなく B を標本空間として出発したときの A の確率を求めている。別の言い方をすれば、 B が保証されているときの A の確率である。

一般の確率論の場合にも、似た公式がある：

$$P(A|B) = \frac{\sum_{\omega \in A \cap B} p(\omega)}{\sum_{\omega \in B} p(\omega)} = \sum_{\omega \in A \cap B} \frac{p(\omega)}{P(B)}, \quad \text{ただし } P(B) \neq 0.$$

この公式はかなり見た目がよくない。今のところは、この公式が存在することを認識しておけば十分である。そしてすぐ後で、明確な例を通してこれを分かりやすくする。

いずれの場合でも、つまり古典的であってもそうでなくても、 $P(A \text{ かつ } B)$ の公式を変形すれば、条件付き確率の別の表現が得られる：

$$P(A \text{ かつ } B) = P(A | B) \cdot P(B) \quad \longrightarrow \quad P(A | B) = \frac{P(A \text{ かつ } B)}{P(B)}, \quad P(B) \neq 0.$$

A と B の役割を入れ替えた場合にも、同じ過程が成り立つ：

$$P(B \text{ かつ } A) = P(B | A) \cdot P(A) \quad \longrightarrow \quad P(B | A) = \frac{P(A \text{ かつ } B)}{P(A)}, \quad P(A) \neq 0.$$

これらの公式は、条件付き確率を求めるための便利な方法を与えてくれる。もし $P(A \text{ かつ } B)$ と、知っているとして仮定している事象の確率の両方が分かっているならば、上の二つの公式のいずれかを用いて条件付き確率を計算できる。これを二つの例で示そう。

1.2 例1：古典的な場合の条件付き確率

第5回講義では、条件付き確率を簡単に見た。そのとき、次の例に触れた。公平な dice を一つ振り、次の二つの事象を考える：

$$A = \{\text{dice が偶数の目を出す}\},$$

および

$$B = \{\text{dice が 4 より小さい目を出す}\}.$$

通常の 6 面 dice を扱っているので、標本空間および事象 A と B は集合記法で次のように書ける：

$$S = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{1, 2, 3\}.$$

今、あなたが dice を投げ、それが空中で回っているとき、時間が一瞬ゆっくりになり、神が天から降りてくると想像しよう。神はあなたの耳元で「dice は 4 より小さい目で止まる」とささやく。この瞬間、あなたは事象 B が起こることを完全に知ることになる。それによって、事象 A が起こることに関する不確実性も変化するかもしれない。

B が保証されているときに A が起こる新しい確率を求めるために、標本空間全体を、4 より小さい数の小さな部分集合へ実質的に縮小することができる：

$$S = \{1, 2, 3, \cancel{4}, \cancel{5}, \cancel{6}\} \quad \longrightarrow \quad \{1, 2, 3\}$$

この制限された体系では、可能な結果が三つあり、そのうち偶数であるものは一つだけであることに気づく。したがって、偶数が出る確率は、最初の $\frac{1}{2}$ から $\frac{1}{3}$ へと変化する。ここで行ったことは、事象 A の条件付き確率を、 B のもとで計算することである。記号で書くと、条件付き確率 $P(A | B)$ は、 A と B が同時に起こる確率の公式を変形することで計算される：

$$P(A \text{ かつ } B) = P(A | B) \cdot P(B) \quad \longrightarrow \quad P(A | B) = \frac{P(A \text{ かつ } B)}{P(B)}.$$

見れば分かるように、集合 A と B の共通部分は $A \cap B = \{2\}$ である。したがって、上で行った議論は記号で次のように書ける：

$$P(A | B) = \frac{|A \cap B|}{|B|} = \frac{|\{2\}|}{|\{1, 2, 3\}|} = \frac{1}{3}.$$

おそらく、まず $P(A \text{ かつ } B)$ と $P(B)$ を直接計算し、それを公式に代入する方が簡単である：

$$P(A | B) = \frac{P(A \text{ かつ } B)}{P(B)} = \frac{1}{6} \div \frac{3}{6} = \frac{1}{3}.$$

1.3 例2：一般の場合の条件付き確率

今度は、この dice を不公平なものに置き換え、次の確率を持つとする：

$$p(1) = 0.1, \quad p(2) = 0.2, \quad p(3) = 0.1, \quad p(4) = 0.1, \quad p(5) = 0.1, \quad p(6) = 0.4.$$

前と同じ事象と標本空間を用いる：

$$S = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{1, 2, 3\}.$$

ここでは古典的確率を扱っていないことに注意しよう。つまり、部分集合の大きさを単に数えるだけでは確率を求めることはできない。代わりに、それぞれの事象に含まれる個別の結果の確率を足し合わせる必要がある。したがって、 A と B の確率は次のようになる：

$$B = \{1, 2, 3\} \implies P(B) = p(1) + p(2) + p(3) = 0.1 + 0.2 + 0.1 = 0.4$$

$$A = \{2, 4, 6\} \implies P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7$$

したがって、この不公平な体系では、dice を投げたときに偶数が出る確率は 70% である。今、再び dice を空中へ投げたとき、神が降りてきて、結果は必ず 4 より小さい数になると教えてくれたとする。この状況では、事象 A に関する不確実性は再び変化する。なぜなら、4、5、または 6 が出る可能性が取り除かれたからである。したがって、与えられた情報に基づいて、新しい体系は縮小された標本空間を持つ：

$$S = \{1, 2, 3, \cancel{4}, \cancel{5}, \cancel{6}\} \longrightarrow \{1, 2, 3\}$$

大まかに言えば、この新しい体系では、結果 1, 2, 3 の確率を、合計が 1 になるように調整しなければならない。数学的には、それぞれを縮小された標本空間の確率 $P(B) = 0.4$ で割ることでこれを行う。実際、そのようにすると、すべての可能な結果の和は確かに 1 になる：

$$\frac{p(1)}{P(B)} + \frac{p(2)}{P(B)} + \frac{p(3)}{P(B)} = \frac{0.1}{0.4} + \frac{0.2}{0.4} + \frac{0.1}{0.4} = \frac{0.4}{0.4} = 1.$$

したがって、 B の中にある個別の結果の条件付き確率は、元の確率を縮小された標本空間全体の確率で割ることで求められる。特に、2 が出る新しい確率を見ると、

$$p(2 | B) = \frac{p(2)}{P(B)} = \frac{0.2}{0.4} = \frac{1}{2}.$$

ここで、 $p(2 | B)$ は $P(\{2\} | B)$ の略記である。ある意味で、これは自然である。元の体系では、三つの結果 1, 2, 3 の確率はそれぞれ 0.1, 0.2, 0.1 であった。1 または 3 が出る確率は等し

く 0.1 であり、2 が出る確率はその二倍の 0.2 である。したがって、この三つの結果だけに突然注目を制限したとしても、この比率は保たれるはずである。つまり、 $P(2 | B)$ は $P(1 | B)$ と $P(3 | B)$ の二倍であり、後者二つは等しいはずである。これを実現する唯一の方法は、次の確率である：

$$P(1 | B) = \frac{1}{4}, \quad P(2 | B) = \frac{1}{2}, \quad \text{および} \quad P(3 | B) = \frac{1}{4}.$$

この体系で条件付き確率 $P(A | B)$ を計算するために、まず $P(A \text{ かつ } B)$ を計算し、それから前と同じ公式を用いることができる。強調しておく、ここでは古典的確率を扱っていないので、複合事象 $A \text{ かつ } B$ の確率を求めるには、共通部分 $A \cap B$ の中を見て、その集合に含まれる結果の確率をそれぞれ足し合わせる必要がある。 $A \cap B = \{2\}$ なので、結果は一つだけ、つまり 2 が出ることであり、したがって和には一つの項しかない。単純に、 $P(A \text{ かつ } B) = 0.2$ と結論できる。

ここで、条件付き確率の公式を用いて、先ほどの議論を確認できる：

$$P(A | B) = \frac{P(A \text{ かつ } B)}{P(B)} = \frac{0.2}{0.4} = 0.5,$$

これは前と同じ答えである。

演習

不公平な dice について、条件付き確率 $P(B | A)$ を計算せよ。

解答

求めたいのは $P(B | A)$ である。つまり、dice が偶数で止まったことをすでに知っているときに、それが 4 より小さい数である確率である。共通部分は

$$A \cap B = \{2\},$$

なので、前と同じく $P(A \text{ かつ } B) = 0.2$ である。ただし、今回は公式を適用する前に $P(A)$ を計算する必要がある。これは

$$P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7,$$

であり、したがって

$$P(B | A) = \frac{P(A \text{ かつ } B)}{P(A)} = \frac{0.2}{0.7} = \frac{2}{7}.$$

1.4 ベイズの定理

前の演習が示すように、二つの条件付き確率

$$P(A | B) \quad \text{および} \quad P(B | A)$$

は一般には同じではない。しかし、それらは互いに関係している。Section 1.1 の議論によれば、確率 $P(A \text{ かつ } B)$ を求める方法は二つある：

$$P(A \text{ かつ } B) = P(A | B) \cdot P(B), \quad \text{および} \quad P(A \text{ かつ } B) = P(B | A) \cdot P(A).$$

この二つを等しいものとして並べ、変形すると、

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

この関係はベイズの定理として知られている。

ベイズの定理

事象 A と B について、 $P(A) \neq 0$ かつ $P(B) \neq 0$ ならば、

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

ベイズの定理は、一方の条件付き確率が分かっていて、もう一方を計算したいときに特に有用である。例えば、前節の不公平な dice の例に基づけば、

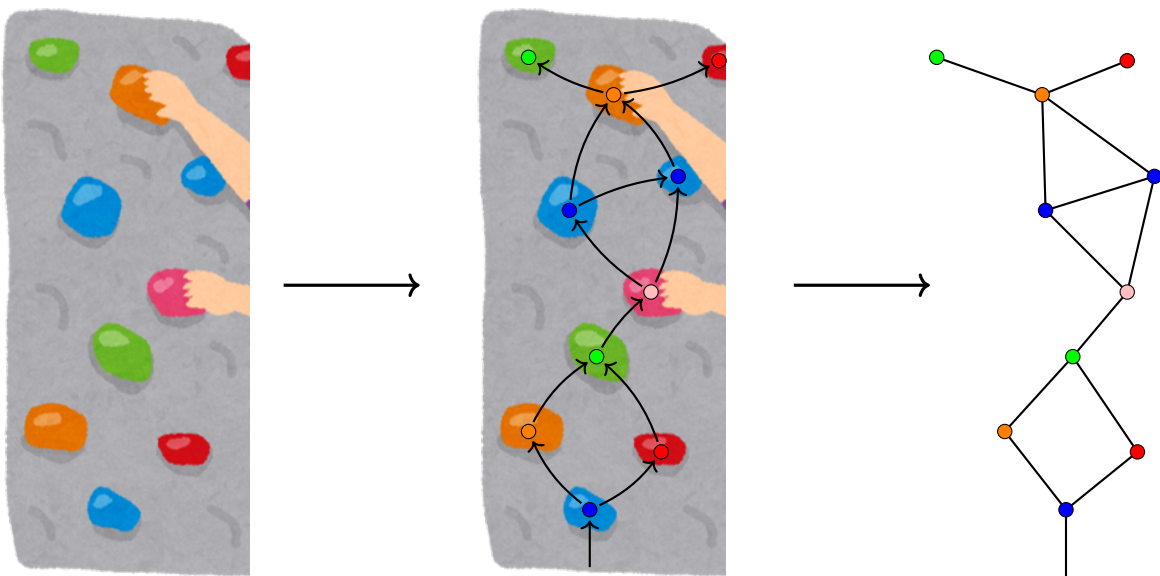
$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A)} = \frac{0.5 \times 0.4}{0.7} = \frac{2}{7}.$$

2 決定木

条件付き確率は、互いに影響し合うかもしれない連続した選択を考えるときに自然に現れる。小さな体系では、起こり得るすべてのことの空間を単純に描き出してみることが非常に役立つ。この方法の一つが決定木である。これは、複数段階の過程の各段階に応じて枝分かれする図である。これを動機づけるために、少し複雑な例から始めよう。

2.1 動機づけの例：ロッククライマー

岩壁を登る人の手を追うとする。この状況は、色付きの岩のネットワークがどのように互につながっているかを表示する一種の「グラフ」として理想化できる：



この図では、クライマーが岩を登るにつれて、その手は一番下の青い岩から上方へ進み、最上部の緑または赤の岩へ至る経路をたどる。この過程のいくつかの段階では、ロッククライマーはどちらへ進むかの選択を与えられていることに注意しよう。例えば、岩壁の最初のところでは、青い岩からオレンジの岩へ進むこともできるし、おそらく赤の岩へ進むこともできる。他の点では、クライマーは何をするか選ぶことができず、それは単に決まっている。例えば、最初のオレンジの岩からは、次の緑の岩へ進まなければならない。

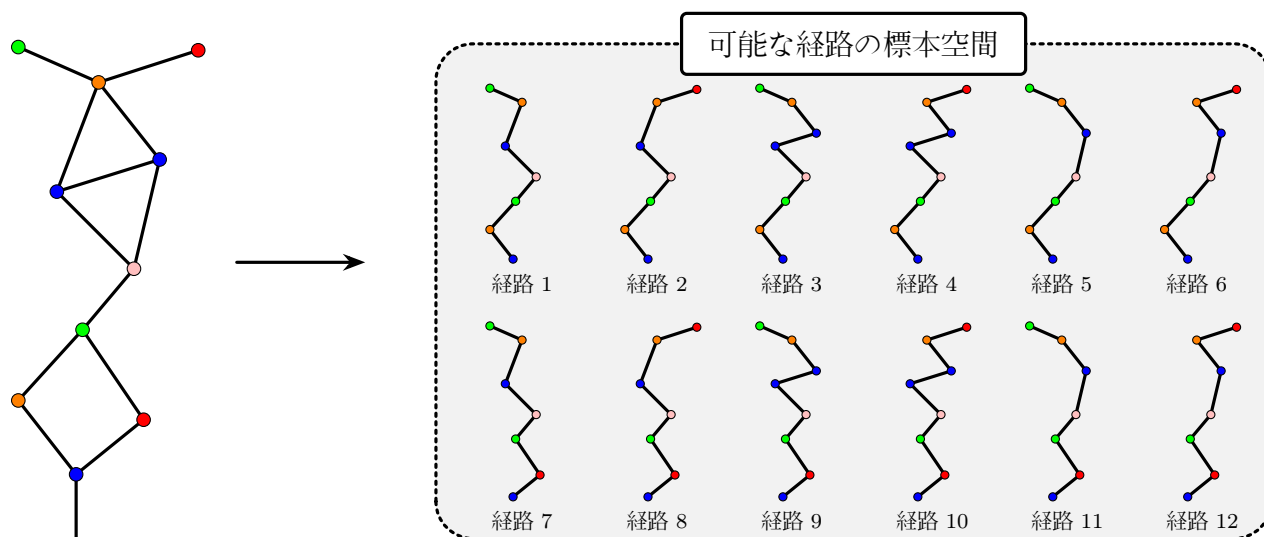
クライマーが選択をするたびに、左の岩を選ぶ確率が 50%、右の岩を選ぶ確率が 50% であるでしょう。中心的な問いは次である：

クライマーがある特定の経路を通る確率はいくらか。

経路が 12 本あると仮定すると、最初の自然な推測は、古典的確率の Golden Formula を適用することである：

$$P(\text{ある特定の経路}) = \frac{1}{\text{経路の総数}}.$$

まず、標本空間を記述する必要がある。この場合、標本空間は、下から上までのすべての経路の空間である。少し時間をかければ、岩壁を通る一意な経路は実際に 12 本あることを納得できるだろう。それらを下に示す：



したがって、この設定における標本空間は次の集合である：

$$S = \{\text{経路 1}, \text{経路 2}, \text{経路 3}, \dots, \text{経路 12}\}.$$

したがって、上の公式を使って特定の経路の確率を計算しようとするのは自然に思える。例えば、経路1について、次のように推測できる：

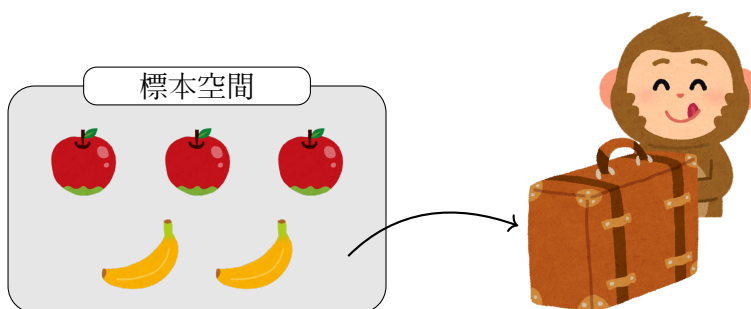
$$P(\text{経路 1}) \stackrel{?}{=} \frac{1}{\text{経路の総数}} = \frac{1}{12}.$$

実のところ、これは正しくない。後でこれを正しく計算するが、今はまず次のことに注意しておこう。経路の空間では、ある経路は他の経路よりも複雑である。これらの異なる経路は同様

に起こりやすい結果ではないので、上の公式は正しい方法ではない。この点にはすぐ戻る。その前に、より単純な設定でもう一度この概念を説明しよう。

2.2 例：ジャングルと泥棒

あなたがジャングルにいて、バッグを放置したとする。地元のサルが木から降りてきて、あなたのバッグをランダムにあさる。たまたま、昼食用に持ってきた果物がバッグに 5 個入っている：3 個のりんごと 2 本のバナナである。サルはランダムに果物を盗み、見つけたものを食べてから、もう一度バッグをのぞく。確率論では、この過程を復元なしの抽出と呼ぶ。



演習 1

- (a) サルが最初にりんごを盗む確率はいくらか。
- (b) サルが最初にバナナを盗む確率はいくらか。

解答

果物は全部で 5 個あるので、単純な数え上げを行えばよい：

- (a) $P(\text{最初にりんご}) = \frac{3}{5}.$
- (b) $P(\text{最初にバナナ}) = \frac{2}{5}.$

次に、連続した問いに注目しよう：

サルがバナナを盗み、その後にりんごを盗む確率はいくらか。

ここでは条件付き確率を用いるべきであることは明らかである。なぜなら、二回目の結果は一回目の結果に明らかに影響されるからである。

順序を正確にするために、 B_1 は「一回目にバナナ」、 A_2 は「二回目にりんご」を意味するとする。すると「バナナ、その後にりんご」は「 B_1 かつ A_2 」という事象である。

これは条件付き確率の公式を用いて計算できる。次の公式を用いればよい：

$$P(B_1 \text{ かつ } A_2) = P(A_2 | B_1) \cdot P(B_1).$$

この公式を一段階ずつ評価しよう：

ステップ 1: 先ほど見たように、 $P(B_1) = \frac{2}{5}$ である。

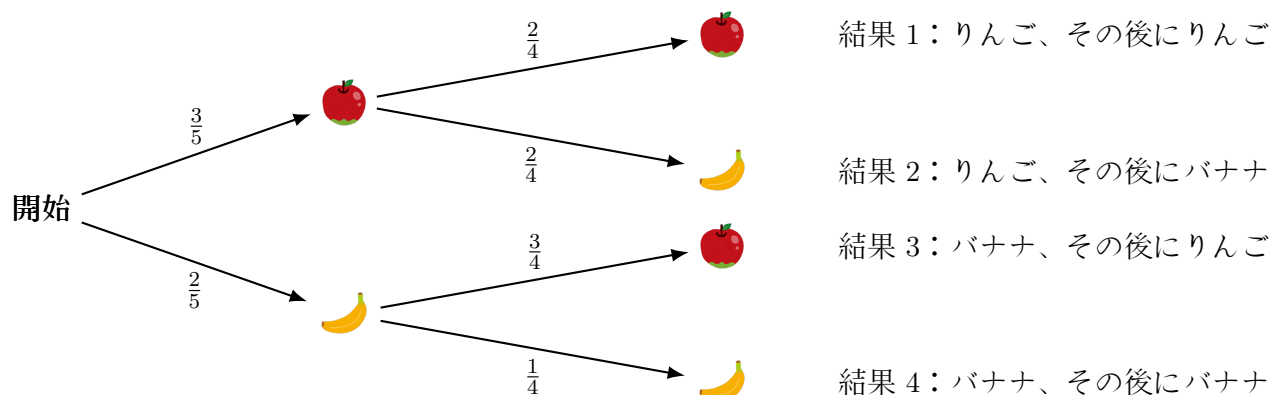
ステップ 2: 次に $P(A_2 | B_1)$ を計算する。バナナが一つ取り除かれ、サルに食べられたとすると、二回目にバッグから果物を取るとき、サルは異なる標本空間を持つことになる。一回目に果物を取るときには、果物は 5 個あった。しかし、バナナがすでに一つ取り除かれているので、残っている果物は 4 個だけである：3 個のりんごと 1 本のバナナである。この新しい標本空間の大きさは 4 であり、サルが選べるりんごは 3 個ある。したがって、条件付き確率は

$$P(A_2 | B_1) = \frac{3}{4}.$$

ステップ 3: これら二つの数を掛け合わせることで、バナナを取り、その後にりんごを取る確率を得る：

$$P(B_1 \text{ かつ } A_2) = P(B_1) P(A_2 | B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}.$$

同様の推論は、このいたずら好きなサルに関する他の可能性にも適用できる。この過程全体を木として整理すると、下のようになる：



見て分かるように、先ほど行った計算、すなわち

$$P(\text{バナナ、その後にりんご}) = P(B_1) \cdot P(A_2 | B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10},$$

は、上の木で結果3に対応する枝を確認することでも視覚的に確かめられる。ここでは、結果3に対応する枝に書かれた数を単に掛け合わせているのである。実際、ここで一般的な技法の最初の例を見たことになる。複数段階の実験の結果を木として整理できるなら、任意の結果の確率は、その枝に沿って現れるすべての確率を掛け合わせることで得られる。

この観察を使って、上の木にある他の三つの結果も素早く計算できる：

- 結果1: $P(A \text{ その後 } A) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$
- 結果2: $P(A \text{ その後 } B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$
- 結果4: $P(B \text{ その後 } B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}.$

これら四つの確率の和は

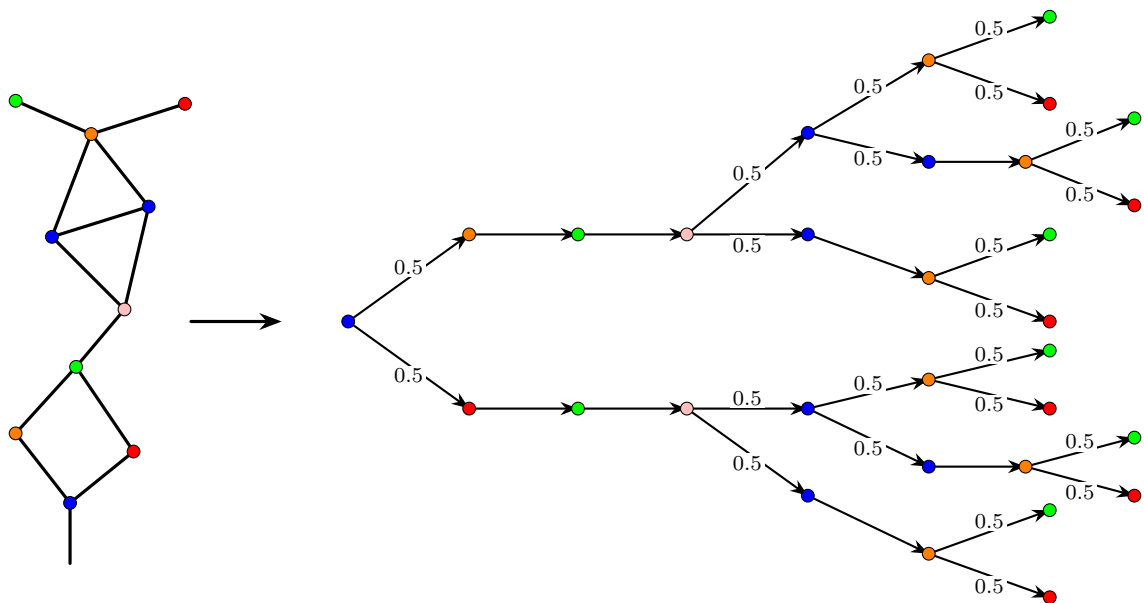
$$\frac{3}{10} + \frac{3}{10} + \frac{3}{10} + \frac{1}{10} = 1.$$

これは、必ずいずれか一つの結果が起こることと同じである。

2.3 ロッククライマーに戻る

ここまでで、決定木を用いれば、複数段階の実験におけるすべての可能な結果を整理できることを見た。特定の経路の確率を求めるには、その経路に沿って書かれている確率を掛け合わせればよい。この技法を用いて、ロッククライマーが壁を登るそれぞれの経路に対応する確率を求めることができる。

まず、この体系の木を書き、それぞれの「選択点」に仮定された 50-50 の確率を付ける。完全な木は下に描かれている。



上の木は、それぞれの経路が順番に並ぶように整理されている。最初の経路が経路1に対応し、二番目の経路が経路2に対応し、以下同様である。ここで、ある経路は他の経路よりも複雑であることに注意しよう。例えば、経路12は経路9や10よりも単純に見える。これは、壁を登る経路の中には、三回の選択だけで済むものもあれば、四回の選択が必要なものもあるからである。実際、ピンクの岩から左へ進む選択を含む経路は、代わりに右へ進む選択を含む経路よりも複雑になる。いずれにせよ、決定木の理論を用いれば、岩壁を登る任意の経路の確率を計算できる。klepto-monkey の例と同様に、この場合も、その枝に書かれている値を掛け合わせれば全体の確率が得られる。結果を下にまとめる。

各経路の確率

各経路が必要とする選択の回数によって区別する：

- 経路 1, 2, 3, 4, 7, 8, 9 および 10 は、すべて四回の選択を必要とする。決定木では、これらの経路に対応する枝に 0.5 が 4 個書かれているので、それらの全体の確率は

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}.$$

- 経路 5, 6, 11 および 12 は、合計で三回の選択だけを必要とする。したがって、これらの経路のそれぞれを取る確率は、他のものよりも大きい：

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

計算が正しいかどうかは、すべての答えを足して 1 になることを確認すれば、いつでも二重にチェックできる。実際、

$$8 \left(\frac{1}{16} \right) + 4 \left(\frac{1}{8} \right) = \frac{1}{2} + \frac{1}{2} = 1,$$

となり、これは正しい。ここで、もう一つのことも分かった。クライマーが単純に見える経路の一つを取る確率は 50% であり、そうでない確率も 50% である。

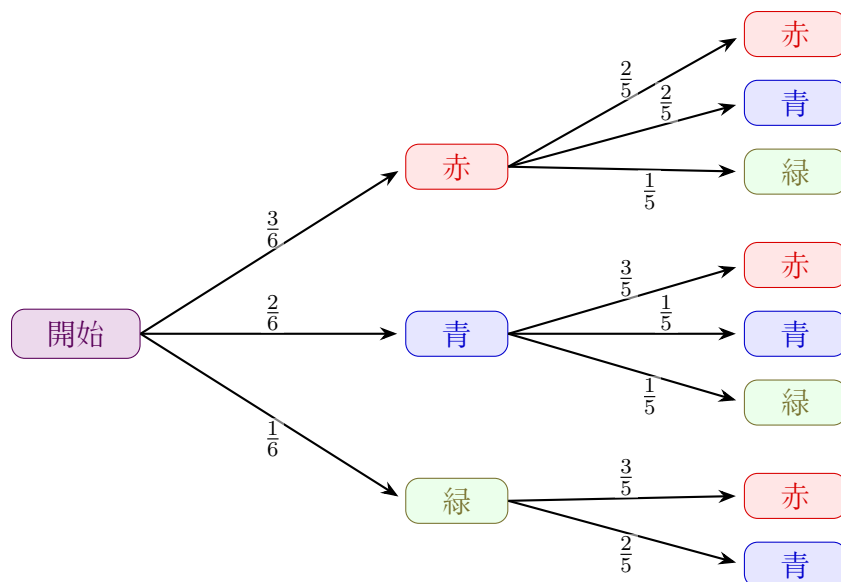
2.4 演習：復元なしの抽出

箱の中に 6 個の玉が入っている：3 個の赤、2 個の青、1 個の緑である。最初の玉を箱に戻さずに、ランダムに 2 個の玉を取る。

演習 2

- (a) 赤を二回連続で取る確率はいくらか。
- (b) 緑、その後に青を取る確率はいくらか。
- (c) 青と赤の玉を、どちらの順序でもよいとして取る確率はいくらか。

完全な解答に入る前に、木がどのように見えるかを観察しよう。完全性のために、それぞれの矢印に沿ってすべての確率を記入してある：



見て分かるように、下の枝では、もはや三つの可能な結果は存在しない。なぜだろうか。それは、箱の中には緑の玉が 1 個しかないからである。もしそれを最初を選ぶと、箱の中には赤と青の玉しか残らない。この木には、すべての確率と条件付き確率が完全に記入されている。これらを公式で直接計算する代わりに、Section 2.2 のサル泥棒の場合と似たやり方で、発見的に議論できる。例えば、赤を引いた直後にもう一つ赤を引く確率を計算するには、赤が一つすでに取り除かれた状況を見ればよい。このとき標本空間は、残っている 2 個の赤、2 個の青、1 個の緑だけに縮小されている。5 個のうち選べる赤の玉は 2 個なので、条件付き確率 $P(\text{赤} | \text{赤})$ は $\frac{2}{5}$ である。これを踏まえて、上の問いに答えよう。

解答

(a) 赤、その後に赤：

$$P(\text{赤、その後に赤}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{6}{30} = \frac{1}{5}.$$

(b) 緑、その後に青：

$$P(\text{緑、その後に青}) = \frac{1}{6} \cdot \frac{2}{5} = \frac{2}{30} = \frac{1}{15}.$$

(c) 青と赤とは、赤その後に青、または青その後に赤を意味する：

$$P(\text{赤、その後に青}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5},$$

また、

$$P(\text{青、その後に赤}) = \frac{2}{6} \cdot \frac{3}{5} = \frac{1}{5}.$$

これらは互いに排反な結果なので、

$$P(\text{青と赤}) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.$$

3 有名な例

3.1 モンティ・ホール問題

あなたがゲーム番組に出ているとする。三つの扉がある。一つの扉の後ろには車があり、他の二つの扉の後ろにはヤギがいる。扉はランダムに配置されているので、どの扉の後ろに何があるかは分からない。そこで、あなたは車があることを願って扉1を選ぶ。扉の後ろを知っている司会者は、別の扉を開け、ヤギを見せる。司会者は常に、ヤギがいる別の扉を開け、常に切り替える機会を与えるものと仮定する。このとき、あなたには選択肢がある：

最初に選んだ扉にとどまる または 残っている開いていない扉へ切り替える。

切り替えるべきだろうか。

驚くべきことに、これらの標準的な仮定のもとでは、答えは yes である。切り替えた方が理にかなっている。最初の扉にとどまって車を当てる確率は $\frac{1}{3}$ であり、扉を切り替えて車を当てる確率は $\frac{2}{3}$ である。これはかなり分かりにくいので、三つの異なる方法で議論しよう。

3.1.1 議論1：確率を移す

$$A = \{\text{最初に選んだ扉の後ろに車がある}\}$$

とする。このとき、

$$P(A) = \frac{1}{3},$$

したがって、

$$P(A \text{でない}) = \frac{2}{3}.$$

最初の選択が正しかったなら、切り替えると負ける。最初の選択が間違っていたなら、司会者がもう一方のヤギを取り除くので、切り替えると勝つ。したがって、

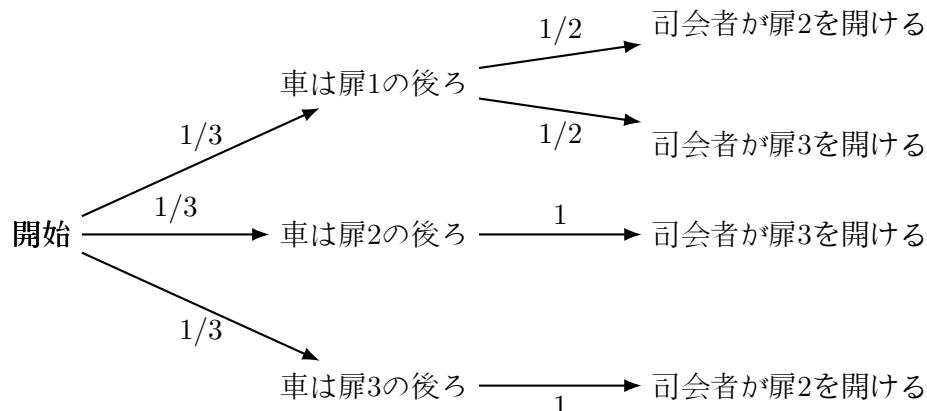
$$P(\text{切り替えると勝つ}) = P(\text{最初の選択が間違っていた}) = \frac{2}{3}.$$

3.1.2 議論2：ゲームを大きくする

代わりに、扉が 1000 個あり、そのうち一つの扉の後ろには車があり、残り 999 個の扉の後ろにはヤギがいるとしよう。前と同じように、あなたは扉1を選んだとする。車が実際にどこにあるかを知っている司会者は、998 個の扉を開け、その後ろにヤギがいることを示す。残った扉が二つ、つまりあなたの扉ともう一つだけになったとき、あなたは最初の選択を疑い始めるかもしれない。もし扉が 1000 個あったなら、あなたの最初の選択が正しい確率は $\frac{1}{1000}$ である。その確率は、司会者がすでにヤギだと知っていた扉を開けたからといって突然大きくなるわけではない。むしろ、あなたの扉が車を含む確率は同じ $\frac{1}{1000}$ のままであり、もう一つ残った扉に車がある可能性が圧倒的に高い。明らかに、扉を切り替える方がよい考えである。

3.1.3 議論3：決定木

最初に扉1を選んだと仮定すると、可能な結果の空間は下のようなになる。ここで強調すべきことは、司会者は車が実際にどの扉の後ろにあるかを知っているということである。ヤギを見せる扉を意図的に選びたいなら、司会者には選択肢がある場合もあれば、ない場合もある。司会者が二つのヤギの扉を選べる場合には、それらを等確率で選ぶと仮定する。



もし車が扉1の後ろにあったなら、切り替えると負ける。上の決定木によれば、これが起こる確率は上の二つの枝の確率の和であり、結果は $\frac{1}{3}$ である。車が扉1の後ろにないなら、それは扉2または扉3の後ろにある。この確率は $\frac{2}{3}$ である。したがって、切り替えれば $\frac{2}{3}$ の確率で車を当て、切り替えることで $\frac{1}{3}$ の確率で車を失うことが分かる。

3.2 誕生日のパラドックス

A を「部屋の中の少なくとも二人が同じ誕生日を持つ」という事象とする。この場合、確率 $P(A)$ は、部屋の中の人数 n によって大きく変化する。例えば、 $n=2$ なら、この二人が同じ誕生日を持つ可能性は非常に低いと予想される。しかし、 $n > 365$ なら、うるう年を無視すれば誕生日は 365 通りしかないので、部屋の中の二人が同じ誕生日を持つことは保証される。このように、 n が非常に小さいとき $P(A)$ も小さく、 n が大きくなると $P(A)$ は最終的に 1 になる。ここで問いがある：

どの n の値で、 $P(A)$ は 50% を超えるのだろうか。

別の言い方をすれば、部屋に何人いれば、二人が同じ誕生日を持つことの方が起こりやすいと言えるのだろうか。実は、その答えは非常に驚くべきものである。必要なのはたった 23 人である。

この驚くべき事実は「誕生日のパラドックス」として知られている。これは「真実らしいパラドックス」の一例である。つまり、数学的には正しいが、あまりに意外なので間違っているように感じられる結果である。しかし、これから示すように、誕生日のパラドックスは確率の基本的な応用から従う。

3.2.1 誕生日のパラドックスの解決

誕生日のパラドックスを解決するために、問題を逆向きに言い換える。 $P(A)$ を計算する代わりに、 $P(A\text{でない})$ 、つまり A の補事象の確率を計算する。補事象は、確率論における「ない」という言葉の対応物であることを思い出そう。ここでは、補事象 A^c は「部屋の中の全員の誕生日がすべて異なる」という意味である。

補事象の確率は単に $1 - P(A)$ であることを思い出そう。したがって、もし $P(A\text{でない}) < 50\%$ なら、

$$P(A) + P(A\text{でない}) = 1$$

なので、 $P(A) > 50\%$ である。各誕生日の選択は独立であり、かつ同様に起こりやすいと仮定しよう。つまり、任意の誕生日の確率は $1/365$ である。これから、小さい n について $P(A\text{でない})$ を計算し、過程がどのように動くかを見る。

$n = 2$ の場合、二人の誕生日が異なる確率を計算しようとしている。明らかに、Person 1 の誕生日は自由に選んでよい。しかし、二人の誕生日が異なるためには、Person 2 の誕生日は最初の選択を避けなければならない。Person 1 の誕生日と異なる日は 364 日あるので、ランダムな事象として扱えば、Person 2 が異なる誕生日を持つ確率は

$$P(A\text{でない}) = \frac{\text{Person 1 の誕生日と異なる日の数}}{\text{一年の日数}} = \frac{364}{365} \approx 0.997 = 99.7\%.$$

次に、部屋に三人目を加える、すなわち $n = 3$ とする。二人の誕生日が異なる確率は $\frac{364}{365}$ であることをすでに確認した。しかし、三人目の誕生日も Person 1 の誕生日かつ Person 2 の誕生日と異なっている必要がある。次の人の誕生日に関する確率は、前の人たちの誕生日に条件付けられている。このような「かつ」の文の確率を計算するには、対応する確率を掛け合わせる必要がある：

$$\begin{aligned} P(A\text{でない}) &= \frac{364}{365} \times \frac{\text{Person 1 と Person 2 の誕生日と異なる日の数}}{\text{一年の日数}} \\ &= \frac{364}{365} \cdot \frac{363}{365} \approx 0.992 = 99.2\%. \end{aligned}$$

このパターンは、より大きな n についても繰り返せる。 n の値を 1 増やすたびに、次の人が選べる誕生日をどんどん「使い切って」いくことになる。そのため、各段階でますます小さい数を掛ける必要がある。例えば：

$$n = 4 \text{ の場合： } P(A\text{でない}) = \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \approx 0.983 = 98.3\%$$

$$n = 5 \text{ の場合： } P(A\text{でない}) = \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \left(\frac{361}{365}\right) \approx 0.973 = 97.3\%$$

$$n = 6 \text{ の場合： } P(A\text{でない}) = \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \left(\frac{361}{365}\right) \left(\frac{360}{365}\right) \approx 0.960 = 96.0\%$$

一般の n について、次の公式が得られる：

$$P(A \text{でない}) = \left(\frac{365}{365}\right) \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{365 - (n-1)}{365}\right).$$

このパラドックスは、確率 $P(A \text{でない})$ が 50% を下回る最小の n を見つけることで解決される。なぜなら、それは $P(A)$ が 50% を上回ることを意味するからである。この公式をより大きな n に対して使っていくと、 $n = 23$ のとき、

$$P(A \text{でない}) = \left(\frac{365}{365}\right) \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{343}{365}\right) \approx 0.493 < 0.5.$$

比較として、 $n = 22$ のときには $P(A \text{でない}) \approx 0.524 > 0.5$ である。したがって、 $n = 23$ のとき $P(A) > 50\%$ となる。

3.3 医療検査

ある人口の 1% が病気を持っているとする。この特定の病気に対する医療検査があり、それについて次が成り立つ：

- もし人が病気を持っていれば、その検査は 99% の確率でこれを検出する。
- もし人がその病気を持っていないければ、その検査は 5% の確率で偽陽性を出す。

あなたが検査を受け、陽性になったとする。このとき、あなたが病気を持っている確率はいくらだろうか。驚くべきことに、答えは約 16.7% である。これをこれから示す。

まず、上の条件を条件付き確率を用いて書き直そう。

次のようにおく：

$$D = \{\text{その人は病気を持っている}\},$$

および

$$T = \{\text{その人は陽性になる}\}.$$

問題が尋ねているのは、条件付き確率

$$P(D | T),$$

すなわち、陽性になった後に実際に病気を持っている確率である。これは逆向きの確率とは同じではないことに注意しよう。 $P(T | D)$ は、その人が実際に病気を持っているときに陽性になる確率を計算している。問題の設定から、

$$P(T | D) = 0.99, \quad P(D) = 0.01, \quad P(D^c) = 0.99, \quad \text{および} \quad P(T | D^c) = 0.05.$$

この情報をベイズの定理と組み合わせれば、欠けている条件付き確率を計算できる：

$$P(D | T) = \frac{P(T | D)P(D)}{P(T)}.$$

これを行うためには、まず $P(T)$ 、つまり陽性になる全体の確率を計算する必要がある。

陽性になる方法は二つある：

- その人が病気を持っていて、正しく陽性になる。
- その人が病気を持っていないが、誤って陽性になる。

これらは互いに排反なので、事象 T を「または」の文として解釈し、その確率を次のように計算できる：

$$P(T) = P(T \mid D)P(D) + P(T \mid D^c)P(D^c) = (0.99)(0.01) + (0.05)(0.99) = 0.0594.$$

これをベイズの定理に代入すると、悪名高い 16.7% が得られる：

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T)} = \frac{0.99 \times 0.01}{0.0594} = \frac{1}{6} \approx 0.167.$$

なぜこれが驚くべきことに感じられるのか

この検査は、誰かが実際に病気を持っているときには、それを検出する能力が非常に高い。しかし、この病気はまれであり、偽陽性の数が比較的多い。これらの偽陽性が、一回の陽性結果の信頼性を下げるため、このような驚くべき結果が生じる。