3 囚人問題再考

式　見　拓　仙

Abstract

This article will investigate the way we can naturally interpret the problem of the three prisoners especially when the prior probabilities that they will be set free are not uniform. Suppose that three prisoners *a*, *b* and *c* are in prison. The prisoner *a* knows that two of them will be executed and one released from prison, and then asks the prison guard 'Even if you will tell me which of the two prisoners, *b* and *c*, is going to be executed, you will give me no information about the probability of my release, since either *b* or *c* is certain to be executed.' The prison guard honestly tells *a* that *b* is to be executed. In naive settings, by a simple calculation, the posterior probability that *a* will be set free remain the same, and hence the *a*'s assertion is justified. It is known, however, that there can be the case that the posterior probability that *a* will be released becomes smaller than its prior probability. This case may seem to be intuitively unacceptable. We will study how this situation can be interpreted.

Keywords: the problem of three prisoners, Bayes' theorem.

1．はじめに

３囚人問題とは，計算によって得られた確率評価と素朴な推論に基づく確率評価との間に食い違いが生じるとして例として有名な問題である．多少の

表現の異同はあるものの，おおむね次のような問題である[1]．3人の囚人a，b，cのうち2人が処刑され，残りの1人が釈放されることをaは知っているが，3人のうち誰が釈放されるのかまったくわからないため，aは各a，b，cの各々が釈放される確率は等しく1/3であると評価している．aが看守に対し「bとcのうち少なくとも1人が処刑されることは確実であるから，その2人のうち処刑される囚人の名前を教えてくれたとしても，私の釈放についての情報をあたえることにはならない」と主張した．看守はその主張に納得し，「bが処刑される」と正直に答えた．また，aが釈放されることになっている場合，看守が「bが処刑される」と答える確率をaは1/2と判断している．aが釈放される確率を求めよ．

正解は1/3である．これはBayesの定理より簡単に計算できる．かくして，aの主張通り，釈放される確率に変化はないのであるという直観が正当化される．しかし，この問題の事前確率を変更すると，aが釈放される事後確率は事前確率に比べて増えることもあれば，減少することもある．aと同じ解釈は通用しなくなる．特に，囚人が2人に減ったにもかかわらず，減少するという結果を直観的に納得することに抵抗が生じる．本稿は，これらの結果に対して，少なくとも不自然さを感じないような納得の仕方を探ることを目的としている．

2．3囚人問題

上記のオリジナルの3囚人問題の正解は1/3であると述べたが，一方で，釈放される囚人は3人からaとcの2人になったから，aが釈放される確率は1/2であるという推論も一見説得力を持っている．直観的な推論とBayes解との間に食い違いが生じる一例である．これらの結果を観察するために，

1 Lindley(1985, pp.41-3), Mosteller(1987, pp.28-9)または市川(1998)などを参照．なお，ゲーム理論の「囚人のジレンマ」とは無関係である．

いくつかの記号を定めておく．A を「a が釈放される」という仮説，B を「b が釈放される」という仮説，C を「c が釈放される」という仮説とする．さらに，a の上記の問いに対し，看守が「b (c) が処刑されるという返答する」という事象を E_b (E_c) とする．求める確率は $\mathbf{P}(A|E_b)$ である．仮定より，

$$\mathbf{P}(A)=\mathbf{P}(B)=\mathbf{P}(C)=\frac{1}{3}. \tag{1}$$

看守は正直であるとの仮定より，データ E_b に対する尤度に関しては，

$$\mathbf{P}(E_b|B)=0,\ \ \mathbf{P}(E_b|C)=1$$

である．a が釈放されるとき，看守が b と答えるか，c と答えるか迷って，b と答える確率は 1/2 と仮定した．すなわち，

$$\mathbf{P}(E_b|A)=\frac{1}{2}. \tag{2}$$

以上の事前分布と尤度情報から事後確率を計算すると，

$$\mathbf{P}(A|E_b)=\frac{\mathbf{P}(E_b|A)\mathbf{P}(A)}{\mathbf{P}(E_b|A)\mathbf{P}(A)+\mathbf{P}(E_b|B)\mathbf{P}(B)+\mathbf{P}(E_b|C)\mathbf{P}(C)}$$

$$=\frac{(1/2)(1/3)}{(1/2)(1/3)+(1/3)}=\frac{1}{3}$$

を得る．したがって，また $\mathbf{P}(C|E_b)=2/3$ となる．事前分布が一様であることから，事後確率は各尤度にのみ比例して決まる．

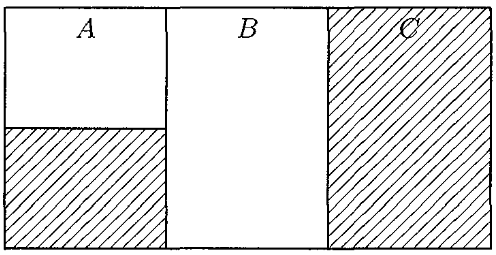
a が釈放される事後確率は $\mathbf{P}(A|E_b)=1/2$ であるという結論にいたる推論として，以下の 2 つのものが考えられる．

(i) 釈放される囚人は a と c の 2 人になったのであるから，2 人で半分ずつになる．

(ii) 釈放される囚人 a と c に対して，a と c の事前確率に比例して，事後の釈放される確率が配分される．

これらのうち (i) の考え方はいささか素朴にすぎるであろう．事前確率を(1)の値ではなく，極端に偏らせた場合も（例えば，$\mathbf{P}(A)=1/1000$，$\mathbf{P}(B)=499/1000$，$\mathbf{P}(C)=500/1000$）(i) のように考えてよいかと問われれば，(i)の推論方法のおかしさに気づくであろう．(ii)の推論も，Bayes 解によれば誤りであるのだが，どのように誤っているのかを，即座に理解するのは困難かもしれない．しかし，容易に観察できるように，事後確率 $\mathbf{P}(A|E_b)$ と $\mathbf{P}(C|E_b)$ が事前確率 $\mathbf{P}(A)$ と $\mathbf{P}(C)$ に比例することと $\mathbf{P}(E_b|A)=\mathbf{P}(E_b|C)$ (= 1)は同値である．したがって，尤度 $\mathbf{P}(E_b|A)=1/2$ が無視されている．少なくとも，A の事後確率を評価するさいに，この尤度情報を適切に組み入れていないということになる．そして，$\mathbf{P}(E_b|A)=1/2$ を無視して，$\mathbf{P}(E_b|A)=1$ と誤認したのはなぜかと問えば，E_b と「b は処刑される」という仮説 B^c とを区別していないことが理由であろうと考えられる．a が釈放されると仮定すれば，b は確実に処刑される．したがって，$\mathbf{P}(B^c|A)=1$ となる．

さらに追求して，E_b と B^c の混同が生じる根拠と探れば，a が看守にどのような質問をしたのかを把握し切れていないことに起因する．a の「b と c のうち処刑されるのはどちらか」という質問に対し，看守が「b である」と答えることが事象 E_b である．よって，単に「b が処刑される」という事象 B^c とは同じではない．a が釈放されるとき，b は c とともに処刑されるが，a への返答に b が選択されなければならない必然性はない．「E_b ならば B^c」は成り立つが，その逆は成り立たないことに注意しなければならない（図 1 参照）.



$\mathbf{P}(A)=\mathbf{P}(B)=\mathbf{P}(C)=1/3$，$\mathbf{P}(E_b|A)=1/2$.

図 1：斜線部分が事象 E_b.

このように，確率空間の構成の誤りに起因するという説明が最も一般的である(例えば，Carlton，2005).

一方，Bayesの定理により得られた$\mathbf{P}(A|E_b)=1/3$を直観的に納得する理路として，「bとcのうち少なくとも1人が処刑されることは確実であるから，その2人のうち処刑される囚人の名前を教えてくれたとしても，私の釈放についての情報をあたえることにはならない」という解釈は，少なくともこの問題に対しては機能していると思われる.

3．事前確率を変更する

事前確率を(1)，尤度を(2)のように決める場合，すでに述べたように，aが釈放される確率は事前と事後に変化はない．しかし，事前確率の設定の設定によっては，aが釈放される確率は変化する．例えば，次の2つの例を見てみよう.

例1． 事前分布は

$$\mathbf{P}(A)=\frac{1}{6},\ \mathbf{P}(B)=\frac{2}{6},\ \mathbf{P}(C)=\frac{3}{6},$$

$\mathbf{P}(E_b|A)=1/2$であるとすると，事後確率は

$$\mathbf{P}(A|E_b)=\frac{(1/6)(1/2)}{(1/6)(1/2)+3/6}=\frac{1}{7},$$

$$\mathbf{P}(C|E_b)=\frac{6}{7}$$

となる.

例2． 事前分布は

$$\mathbf{P}(A)=\frac{1}{6},\ \mathbf{P}(B)=\frac{3}{6},\ \mathbf{P}(C)=\frac{2}{6},$$

$\mathbf{P}(E_b|A)=1/2$であるとすると，事後確率は

$$\mathbf{P}(A|E_b)=\frac{(1/6)(1/2)}{(1/6)(1/2)+(2/6)}=\frac{1}{5},$$

$$\mathbf{P}(C|E_b)=\frac{4}{5}$$

となる．

前者の例では a の釈放される確率は事後に減少し，後者の例では増加している．これらの結果は，「b と c のうち少なくとも1人が処刑されることは確実であるから，その2人のうち処刑される囚人の名前を教えてくれたとしても，私の釈放についての情報をあたえることにはならない」という a の主張が事前確率に依存する限定的な納得の方途であって，汎用性にかけることを示している．

では，「囚人 a の釈放に関してなんら情報をもたらさない」という解釈を否定する，これらの例そのものは直観的に納得がいくであろうか．

市川(1998)では，事前分布に一様性を仮定しない3囚人問題を「変形3囚人問題」とよび，数理的な側面と心理学的な側面の双方から詳細に考察している．同書では，この問題を考案した当初は，例1のように，a の事後確率が事前確率より減少してしまう事態を理論的にはわかっても，直観的にはどうしても納得できなかったと述べている．事前確率に比例して，事後の釈放される確率が a，c に配分されるという推論をすれば，事後確率は事前確率より必ず増大する．しかし，このような推論が誤りであることはわかっている．だからといって，$\mathbf{P}(A|E_b)$ が $\mathbf{P}(A)$ より減少することを直観的に納得させる説明にはなっていない．看守の返事によって，a のライヴァルは c のみになったのであるから，変化しないのであるならまだしも，$\mathbf{P}(A|E_b)$ が $P(A)$ より小さくなるということに不自然さを感じるというわけである．

さらに，同書では次のような結果を指摘している．$\mathbf{P}(A)=\mathbf{P}(B)=1/4$，$\mathbf{P}(C)=1/2$ と定め，尤度が $\mathbf{P}(E_b|A)>2/3$ であると仮定すると，

$$\mathbf{P}(A|E_c)<\mathbf{P}(A),\ \mathbf{P}(A|E_b)>\mathbf{P}(A)$$

という結果を得る．この結果は，c に比べて釈放される事前確率がより小さい b が釈放の候補者として残れば，a にとって不利であり，より大きい c が釈放の候補者として残れば，a にとって好ましい（釈放される確率が高まる），ということを意味している．市川(1998，p.57)は「3囚人問題の徹底的なまでの困難なところ」という[2]．

次節以降で，より一般的に定式化された3囚人問題において，釈放される確率の事前と事後の相違をもたらす条件を分析する．そして，そのことを通じて，A の事後確率に関する反直観的な結果に，少なくとも違和感を感じないような道筋を探求したい．

なお，例1，2の双方で，$\mathbf{P}(C|E_b)>\mathbf{P}(C)$ が成り立っている．実は，後に見るように，これは常に成り立つ結果である．すなわち，データ E_b が得られた後の c の釈放される確率は必ず事前の釈放される確率よりも大きくなる．これは $\mathbf{P}(E_b|C)=1$ であること，すなわち「C ならば E_b」であることによる[3]．「C ならば E_b」となる理由をさらにさかのぼれば，a の看守への質問の仕方が「b と c の2人のうちどちらが処刑されるのか」であることと看守が正直であるという仮定による．

4． 3囚人問題の一般化

あらためて，問題を一般的に定式化しておこう．3人の囚人 a，b，c のう

2　このように，直観的な理解が困難な3囚人問題を視覚的に理解する道具として，「ルーレット表現」が提案されている（詳しくは，市川，1988，1998を参照）．ただし，「ルーレット表現」は視覚的・図式的に把握する道具としては有効といえるが，理論的に納得のいく解釈をこれに求めることは難しいと思われる．

3　仮説 C から E_b が演繹できるとする．観測結果は E_b か E_c のどちらかである．E_c が観測されれば，C は反証されるということになる．一方，E_b が観測されたからといって，C が確証されるわけではない．しかし，これを単に反証することができなかったとみなすのではなく，C の正しさの確率が高まったと考える帰納的推論の方法を仮説演繹法とよぶことがある．仮説演繹法はBayesの定理の極端な応用例とみなすことができる．

ち2人が処刑され，残りの1人が釈放されることをaは知っている．aが看守に対し「bとcのうち少なくとも1人が処刑されることは確実であるから，その2人のうち処刑される囚人の名前を教えてくれたとしても，私が釈放についての情報をあたえることにはならない」と主張した．看守はその主張に納得し，「bが処刑される」と答えた．看守は正直者であると仮定し，aが釈放される確率を求めよ．

各a,b,cが釈放される確率（事前確率）をaは

$$\mathbf{P}(A)=\pi_a,\ \mathbf{P}(B)=\pi_b,\ \mathbf{P}(C)=\pi_c \qquad (\pi_a,\ \pi_b,\ \pi_c>0)$$

と判断している．さらに，aが釈放されると場合，看守が「bが処刑されると答える」確率を

$$\mathbf{P}(E_b|A)=l(a;b)\in(0,\ 1]$$

と表わすことにする（E_bの観測後はこれをAの尤度と呼ぶ）．これ以外の尤度は，aの質問の仕方および看守が正直者であるという設定より，

$$\mathbf{P}(E_b|B)=0,\ \mathbf{P}(E_b|C)=1$$

となる．これらの仮定より，事後確率は

$$\mathbf{P}(A|E_b)=\frac{\pi_a l(a;b)}{\pi_a l(a;b)+\pi_c},$$

$$\mathbf{P}(B|E_b)=0,$$

$$\mathbf{P}(C|E_b)=1-\mathbf{P}(A|E_b)=\frac{\pi_c}{\pi_a l(a;b)+\pi_c}. \tag{3}$$

(3)より，$\mathbf{P}(C|E_b)>\mathbf{P}(C)$であることがわかる．

また，事前，事後を問わず，釈放される確率の大きい（小さい）囚人を「強い」（「弱い」）と表現することにする．

5．事後確率の観察

a が釈放される事後確率の振る舞いについて考える．まず，既に述べたように，この事後確率は事前確率に比べて減少することも，変化しないことも，増加することもありうる．事後確率が事前確率より減少する条件は

$$\mathbf{P}(A|E_b) < \mathbf{P}(A) \Longleftrightarrow \frac{\pi_a l(a;b)}{\pi_a l(a;b)+\pi_c} < \pi_a$$

$$\Longleftrightarrow l(a;b) < \frac{\pi_c}{\pi_b+\pi_c}$$

の最後の不等式のように書き表すことができる．増加することの必要十分条件は $l(a;b) > \pi_c/(\pi_b+\pi_c)$ である．また，変化しないということは等号が成り立つことに他ならない．オリジナルの3囚人問題において，a が釈放される確率が変わらないのは，図1からもわかるように，$l(a;b)=1/2=\pi_c/(\pi_b+\pi_c)$ が成り立っているからである．

$\pi_c/(\pi_b+\pi_c)=\mathbf{P}(E_b|A^c)$ であるから，A の確率が減少するのは，尤度比が

$$\frac{\mathbf{P}(E_b|A)}{\mathbf{P}(E_b|A^c)} < 1 \tag{4}$$

となることと同値である．ただし，(4)は3囚人問題に限らず，一般的に成り立つことである．3囚人問題の固有の状況に分け入ってみよう．

$\mathbf{P}(A|E_b)$ は

$$\mathbf{P}(A|E_b) = \frac{l(a;b)}{l(a;b)+\left(\dfrac{1-\pi_a}{\pi_a}\right)\dfrac{\pi_c}{\pi_b+\pi_c}}$$

と表わすことができ，$\mathbf{P}(A)=\pi_a$ を固定すると，$l(a;b)$ の増加関数であり，$\pi_c/(\pi_b+\pi_c)$ の減少関数である．

$\pi_c/(\pi_b+\pi_c)$ は b または c のどちらか一方が釈放される（すなわち，a が処刑される）という状況で，c が釈放される（条件付）確率である．いわば，b 対 c の「勝負」において，c が「勝つ」確率を表わしている．この確率は b

と c の強弱関係を示す指標となっている（b に対する c の強さを示している）．b との比較で c が強くなれば，$\pi_c/(\pi_b+\pi_c)$ は大きくなり，それにともない $\mathbf{P}(A|E_b)$ も減少する．$l(a;b)$ と π_a を固定して，$\mathbf{P}(A|E_b)$ を $\pi_c/(\pi_b+\pi_c)$ の関数と見ると，そのグラフは以下のようになる．

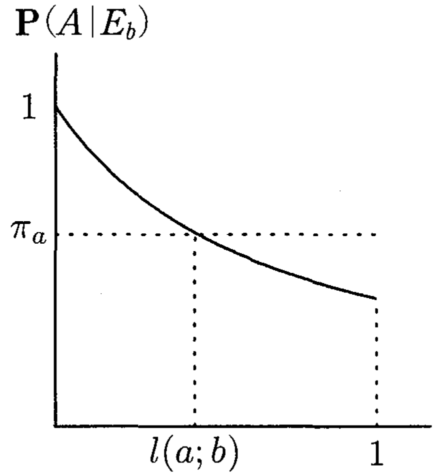


図 2：A の事後確率．

b が処刑されるという看守の返事から，a のライヴァルは c である．$\pi_c/(\pi_b+\pi_c)$ が大きくなるということは，a のライヴァルが強くなるということであるから，それに応じて a が釈放される確率は減少すると解釈できる．そして，$\pi_c/(\pi_b+\pi_c)$ が $l(a;b)$ を凌駕する程度に c が強くなれば，$\mathbf{P}(A|E_b)$ が $\mathbf{P}(A)=\pi_a$ より小さくなるのである．$l(a;b)$ は証拠 E_b がサポートする仮説 A の「尤もらしさ」の程度である．$\pi_c/(\pi_b+\pi_c)$ は b に対する c の相対的な強さの程度である．前者より後者の方が大きくなったとき，$\mathbf{P}(A|E_b)<\mathbf{P}(A)$ となるという十分首肯しうることである．特に，$l(a;b)=1/2$ であるとすれば，$\mathbf{P}(A|E_b)<P(A)$ となる必要十分条件は $\pi_b<\pi_c$，すなわち，b より c の方が相対的に強いことである．$\mathbf{P}(E_b|A)=1/2$ という前提のもとでは，より強い c が釈放の候補者として残れば，a が助かる確率は減少するということである．これも直観的に納得のいくことである．

また，「3 囚人問題の徹底的なまでに困難なところ」についても同様に考

えればよい．もう一度，この問題を述べておく．c が b より強いとし，$\mathbf{P}(E_b|A)>1/2$ であると仮定しよう．このとき，より強い c が釈放の候補者として生き残れば，a にとっては好ましく，より弱い b が生き残れば，a にとって好ましくないという状況が生じるケースがある．確かに，これは直観に著しく反するように見える．しかし，たとえ，c のほうが b より強くとも（すなわち，$\pi_c/(\pi_b+\pi_c)>1/2$ であろうとも），$\pi_c/(\pi_b+\pi_c)$ が $l(a;b)$ を超えない程度の強さならば，$\mathbf{P}(A|E_b)>P(A)$ ともなる（図 3 参照）．いわば，c が b より強くとも，その強さが，a の釈放される確率を侵食するには十分ではないということである．

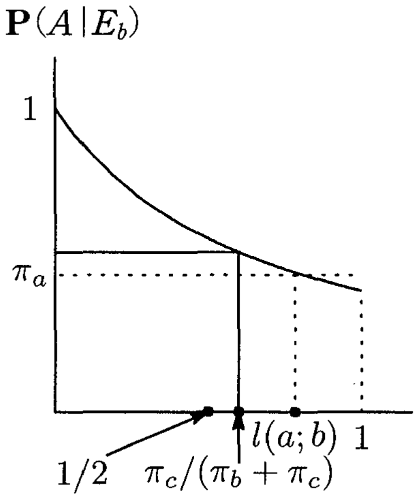


図 3：$\pi_b<\pi_c$ であるとき A の事後確率が増大するケース．

一方，$\pi_c/(\pi_b+\pi_c)<l(a;b)$ は，換言すれば，$\pi_b/(\pi_b+\pi_c)>1-l(a;b)=P(E_c|A)$ である．したがって，看守の返答で E_c が起こる（b が釈放の候補として生き残る）場合，たとえ，b が c より弱くとも，$\pi_b/(\pi_b+\pi_c)>1-l(a;b)$ となる程度の強さをもっていれば，$P(A|E_c)<P(A)$ となる．

参考文献

[1] Carlton, M. A. (2005), Pedigrees, Prizes, and Prisoners: The Misuse of Conditional Probability. *Journal of Statistics Education*, 13 (2)

[2] 市川伸一(1998)『確率の理解を探る』, 共立出版.

[3] 市川伸一(1988)「「納得の道具」としての同型的図式表現」, 数理科学, **26**(3), pp.34-9.

[4] Lindley, D. V. (1985), *Making Decisions* (2nd ed.), John Wiley & Sons.

[5] Mosteller, F. (1987), *Fifty Challenging Problems in Probability with Solutions*, Dover.