■直感が当たらないことだってある !!（条件付確率）

行動経済学という学問分野があります。
人は必ずしも合理的には行動しないということに注目し、伝統的な経済学ではうまく説明できなかった社会現象や経済行動を、人間行動を観察することで実証的にとらえようとする経済学です。

2002 年にダニエル・カーネマンが行動経済学という新しい学問分野を開拓した功績でノーベル経済学賞を受賞しました。同時にバーノン・スミスが実験経済学という学問分野を開拓した功績で受賞。
2013 年にもロバート・シラーという行動経済学者と他２人が資産価格の実証分析に関する功績という理由で受賞。
2017 年にはリチャード・セイラーが行動経済学の理論的発展に貢献したという理由で受賞しています。

そんな行動経済学という新しい学問分野を開拓したダニエル・カーネマンが次のような問題を提供してくれています。

ダニエル・カーネマン『ファスト＆スロー　あなたの意思はどのように決まるか？』
早川書房（上）P295

次の問題をよく読み、質問に直感的に答えてほしい。
[問題]
夜、1 台のタクシーがひき逃げをしました。
この市では、緑タクシーと青タクシーの 2 社が営業しています。
事件とタクシー会社については、次の情報が与えられています。

・市内を走るタクシーの 85%は緑タクシーで、15%が青タクシーである。
・目撃者は、タクシーが青だったと証言している。裁判所は、事件当夜と同じ状況で目撃者の信頼性をテストした結果、この目撃者は青か緑かを 80%の頻度で正しく識別し、20%の頻度でまちがえた。

では、ひき逃げをしたのが青タクシーである確率は何%でしょうか？

この問いでは、直感と計算結果が大きく食い違うのだそうです。
直感では何パーセントくらいだと思いますか？

計算をしてみます。
「タクシーが青だった」という証言があるので、「青いタクシーだという目撃証言が得られたこと」を全事象として（その確率で)、「実際に青のタクシーであった」という事象の確率を割れば、「タクシーが青だった」という条件（情報）の下での条件付確率が求められます。

青を青と認識する（目撃証言する）確率は、$0.15\times0.8=0.12$
緑を青と認識する（目撃証言する）確率は、$0.85\times0.2=0.17$
したがって求める確率は
$0.12 \div (0.12+0.17) \fallingdotseq 0.41$

計算結果は約 41％です。直感では 80%と答える人が多いのだそうです。
みなさんはどうだったでしょうか？

直感と計算結果が大きくズレる理由は、
「本来基準とすべき情報（基準率)」を無視してしまいがちだからだとのことです。
この場合の基準率は「市内のタクシーの 85%は緑タクシーで、15％が青タクシーである」という情報です。
青のタクシーの方が圧倒的に少ないですね、このことをうっかり忘れてしまう。

直感と計算が大きく食い違う可能性があるような問題は大学入試でも時々出題されます。
まず 2018 年の旭川医大の問題を見てみましょう。
腫瘍 X があるかないかを検査したときに、「X があるという診断結果」だったとします。
そのとき「実際に X がある」という確率はどれくらいかというようなことを問う条件付確率の有名問題です。

最初に、直感として〇〇％くらいだという予測をして、その後実際に計算して自分の直感と比較してみるとより深く考えることができると思います。

2018年　旭川医大　医学部

ある臓器にできる腫瘍Xは悪性と良性の2つの型に分けられ、同時に両方の型であることはない。実際にXがある人とない人の割合は3%と97%であり、Xがある人のうち、悪性の人と良性の人の割合は1 : 2である。そして、腫瘍Xがあるかないかを調べる検査Yについて、次の事が知られている。

（ⅰ）悪性のXがある人にYが用いられると、95%の確率でXがあると判定される。

（ⅱ）良性のXがある人にYが用いられると、80%の確率でXがあると判定される。

（ⅲ）Xがない人にYが用いられると、90%の確率でXがないと正しく判定される。

ある人が、この検査Yを受けることになった。このとき、次の確率を求めよ。

（1）　この人にXがあると判定される確率

（2）　Xがあると判定されたとき、悪性のXが実際にある確率

（3）　悪性のXが実際にないとき、Xがないと判定される確率

事象Xを「この人にXがある」

事象X_1を「この人に悪性のXがある」

事象X_2を「この人に良性のXがある」

事象Eを「この人にXがあると判定される」

とします。

（1）　この人にXがあると判定されるのは

「悪性のXがあってXがあると判定される」

「良性のXがあってXがあると判定される」

「XがないのにXがあると判定される」という場合の合計だから

$$P(E)=P(X_1\cap E)+P(X_2\cap E)+P(\bar{X}\cap E)$$

$$=P(X_1)P_{X_1}(E)+P(X_2)P_{X_2}(E)+P(\bar{X})P_{\bar{X}}(E)$$

$$=\frac{3}{100}\times\frac{1}{3}\times\frac{95}{100}+\frac{3}{100}\times\frac{2}{3}\times\frac{80}{100}+\frac{97}{100}\times\frac{10}{100}$$

$$=\frac{1225}{10000}$$

$$=\frac{49}{400}\quad(12.3\%)$$

＊実際に X がある人の割合が 3%であるにもかかわらず、検査を受けると 12.3%の確率で X があると判定されるということですね。
実際には感染していない人が検査を受けて陽性と判定される（偽陽性という）場合があるからです。

（２） X があると判定された（事象 E である）とき、「悪性の X」（X_1）が実際にある確率 $P_E(X_1)$は、

$$P_E(X_1)=\frac{P(X1\cap E)}{P(E)}$$

$$=\frac{95}{10000} \div \frac{1225}{10000}$$

$$=\frac{95}{1225} = \frac{19}{245}$$ （7.8%）

＊X があると判定されても実際に悪性の X があるのは X があると判定された人のうち 7.8%です。
直感との「差」はどうでしたか？

（３） 悪性の X が実際にない場合に X がないと判定される確率はどうなるでしょう。
「悪性の X が実際にない確率」というのは
「実際に X そのものがない確率」＋「良性の X がある確率」

$$= \frac{97}{100} + \frac{3}{100} \times \frac{2}{3} = \frac{99}{100}$$ ・・・①

「そのとき X がないと判定される確率」は
「良性の X があるけれど、X がないと判定される確率」
＋「実際に X がなくて、（正しく）X がないと判定される確率」

$$=\frac{3}{100}\times\frac{2}{3}\times\frac{20}{100} + \frac{97}{100}\times\frac{90}{100} = \frac{877}{1000}$$ ・・・②

となるから、（①の確率を全事象として考えて）②÷①より $\frac{877}{990}$ （88.6%）

＜補足＞このような問題は実戦的には次のような表を作って考えると考えやすいですね。

	X があると判定される	X がないと判定される
悪性の X がある	a	d
良性の X がある	b	e
X がない	c	f

$a = \frac{3}{100} \times \frac{1}{3} \times \frac{95}{100} = \frac{95}{10000}$　　$d = \frac{3}{100} \times \frac{1}{3} \times \frac{5}{100} = \frac{5}{10000}$

$b = \frac{3}{100} \times \frac{2}{3} \times \frac{80}{100} = \frac{160}{10000}$　　$e = \frac{3}{100} \times \frac{2}{3} \times \frac{20}{100} = \frac{40}{10000}$

$c = \frac{97}{100} \times \frac{10}{100} = \frac{970}{10000}$　　$f = \frac{97}{100} \times \frac{90}{100} = \frac{8730}{10000}$

この表より

（１）　この人に X があると判定される確率は

$$a + b + c = \frac{95}{10000} + \frac{160}{10000} + \frac{970}{10000} = \frac{1225}{10000} = \frac{49}{400}$$

（２）　X があると判定されたとき、悪性の X が実際にある確率は

$$\frac{a}{a+b+c} = \frac{95}{10000} \div \frac{1225}{10000} = \frac{19}{245}$$

（３）　悪性の X が実際にないとき、X がないと判定される確率は

$$\frac{e+f}{b+e+c+f} = \frac{40+8730}{10000} \div \frac{160+40+970+8730}{10000} = \frac{877}{990}$$

ちなみに、

「実際に疾患があり、検査の結果、陽性と判定される」ことを・・・・「真陽性(A)」

「実際に疾患があるが、検査の結果、陰性と判定される」ことを・・・「偽陰性(C)」

「実際には疾患はないが、検査の結果、陽性と判定される」ことを・・「偽陽性(B)」

「実際に疾患はなく、検査の結果、陰性と判定される」ことを・・・・「真陰性(D)」

A÷（A＋C）　という比率のことを「感度」

D÷（B＋D）　という比率のことを「特異度」といいます。

「感度」は実際に疾患がある人が検査の結果（正しく）陽性となった人の割合

「特異度」は実際に疾患がない人が検査の結果（正しく）陰性となった人の割合です。

そんなことをテーマにした入試問題をもう１問考えてみてください。

2019 年　成城大　経済学部

ある病原菌の検査試薬は、病原菌に感染している個体に用いると 99%の確率で陽性と判定し､感染していない個体に用いると 99%の確率で陰性と判定する｡ここで､全体の 0.5%がこの病原菌に感染している集団から個体を 1 つ取り出し、この検査試薬を用いて感染しているかどうかを検査する。以下の問いに答えよ。

（1）　この個体が実際に病原菌に感染していて、かつ陽性と判定される確率を求めよ。

（2）　この個体が実際には病原菌に感染しておらず、かつ陽性と判定される確率を求めよ。

（3）　この個体が陽性と判定されたときに、実際には病原菌に感染していない条件付確率を求めよ。

（4）　この個体が陰性と判定されたときに、実際には病原菌に感染している条件付確率を求めよ。

新型コロナウイルス（COVID-19）に対する PCR 検査のことが頭に浮かんでしまうような問題です。しかし、この問題で設定されている数値は実際の新型コロナウイルスに関するものではありません。

（1）　この場合の全事象は､「この集団全体」です。

「この集団全体」で「ある個体が実際に病原菌に感染していて、検査の結果、陽性と判定される」のはどれくらいの割合かを聞かれています。

したがって、全体の 0.5%がこの病原菌に感染していて、そのうち 99%が陽性と判定されるから、

0.5×0.99＝0.495（%）

（2）　この場合も全事象は､「この集団全体」です。

「この集団全体」で「ある個体が実際には病原菌に感染しておらず、検査の結果、陽性と判定される」のはどれくらいの割合かを聞かれています。

したがって、全体の 99.5%がこの病原菌に感染しておらず、そのうちの 1%が陽性と判定されるから、

99.5×0.01＝0.995（%）

＊（1）の数値より多いことに注意。

（3） 検査の結果「陽性」と判定されたけれども、実際には病原菌に感染していない条件付確率を聞かれています。

検査の結果「陽性」と判定されるのは、
「実際に感染していて陽性と判定される ・・・・・・・（1）の場合」と
「実際には感染していないけれど陽性と判定される・・・（2）の場合」の合計だから 0.495＋0.995＝1.49％
これを全事象として考えて、
「病原菌に感染していないときに陽性と判定される確率 0.995％」をこの 1.49％で割って、0.995÷1.49＝$\frac{199}{298}$ （≒66.8％）

＊検査の結果「陽性」と判定された人のうち 66.8％は実際には感染していないという計算結果になりました。直感との「差」はどうでしたか？

（4） 次は、検査の結果「陰性」と判定されたけれども、実際には病原菌に感染している条件付確率を聞かれています。

検査の結果「陰性」と判定されるのは、
「実際に感染していて陰性と判定される場合」0.5×0.01＝0.005（％）と
「実際には感染しておらず陰性と判定される場合」99.5×0.99＝98.505（％）
の合計だから、0.005＋98.505＝98.51％
これを全体事象として考えて、
「実際に感染していて陰性と判定される場合」0.5×0.01＝0.005（％）の比率を求めて

0.005÷98.51＝$\frac{1}{19702}$ （≒0.00508％）

＊検査をして「陰性」と判定された人が「実は感染していました」という可能性は極めて低いという結果になっています。