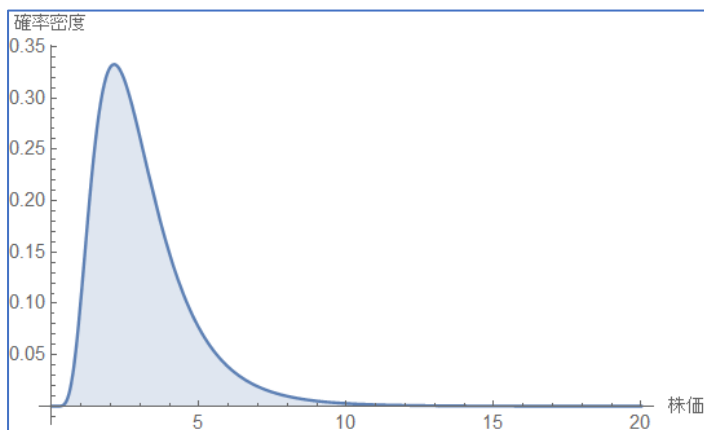
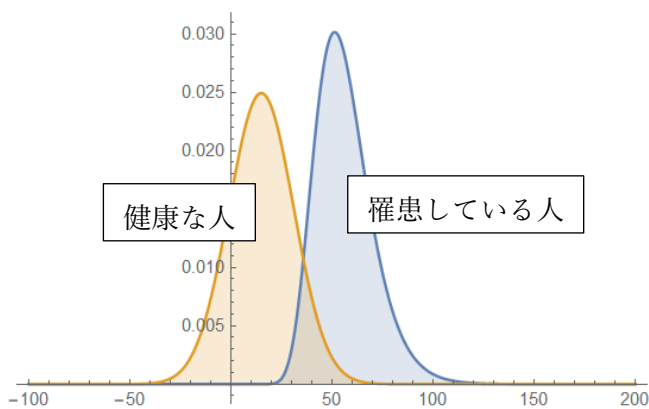


確からしさを表す副詞として、perhaps, probably, surely, absolutely 等があるが、正確さを期するのであれば確率分布を示すのがベストと考える。会話中、「この宝くじの賞金の確率分布はこのように近似できると思います」とディスプレイして見せたり、受験生に「今のままですと、次回の模試の点数の確率分布はこうです」、あるいは、投資家に「この株を買った場合、1週間後の価格はこのような確率分布になります」と提示したり。なぜ、確率分布を示した方が議論が正確になるかというと、同じ期待値(平均値)でも、リスク(分散、あるいは標準偏差)が大きい場合と小さい場合等があり、その様子が一目でわかるから



である。期待値を比較するだけではリスクに関する議論はできない。SF小説で宇宙人が地球人と全く異なる言語体系で意志疎通をするという設定がよくあるが、確率分布で話す宇宙人がいたとしたら、その表現方法はどのようなものであろうか。ビジュアルな確率分布を、テキスト表現することはかなり難しい。暇なので、あれこれ考えてみたがうまくいかない。ビジュアルに「これがその確率分布です」とテレパシーで伝えるほうが、絶対的に情報

量が多くて正確に伝わる。時系列変化も含めるともっと情報量が増える。学生の諸君には、意志決定の際には確率分布で物と考える人になってもらいたいと思う。



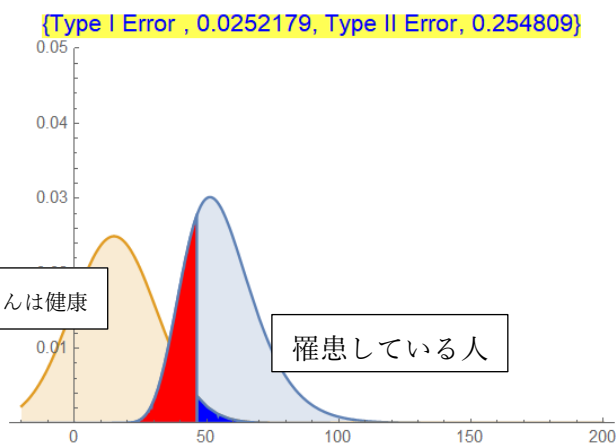
さて、仮説検定における第1種の誤りと第2種の誤りの話をしよう。

現在、新型コロナウイルス感染を疑ったらまずはPCR検査を受けて罹患(りかん)しているか調べましょう、と言われてている。ここでは、動物村のコロナ病(名前は適当につけました)について、抗体検査を行い、コロナ病に感染しているか否かを検定すると仮定する。この抗体検査は検査値が大きいほ

が
罹患率が高いとする。図は、オレンジが健康な人の検査値の確率密度である。青が罹患している人の検査値の確率密度である。

問題：白熊エマさんの検査値は50でした。彼女は罹患していると言えるでしょうか？

問題は、2つの確率密度が重なっている点である。両者が離れていれば、問題はない。しかし、現実には、このように



帰無仮説と対立仮説の確率分布が重なる場合も多い。

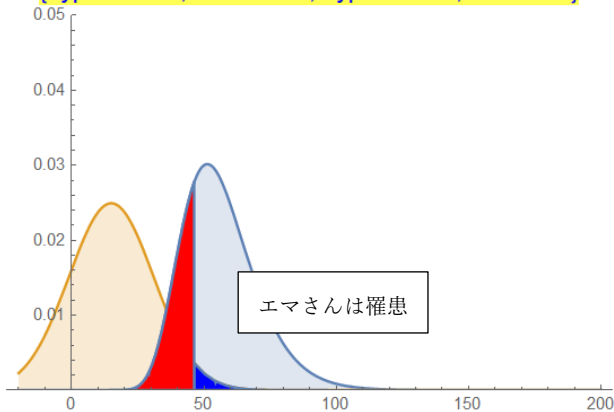
両側検定、有意水準 $\alpha = 5\%$ として、以下のように帰無仮説を立ててみる。

帰無仮説：エマさんは健康である。

対立仮説：エマさんは罹患している。

仮説検定に使う確率分布は、健康な人の確率分布（オレンジの方）を使う。両側検定で、片側で 2.5% の限界値は約 46.3 であった。エマさんの値 50 は、それを超えているので、帰無仮説は棄却されて、罹患している、になる。エマさんが本当に罹患しているのであれば、これは正しい決定であるが。しかし、もしエマさんが健康であった場合、誤りとなる。その場合、帰無仮説が正しかったのに棄却されてしまったことになる。これを**第 1 種の誤り**と呼ぶ。第 1 種の誤りを犯す確率は $\alpha = 5\%$ となる。図中、濃い青の面積が約 2.5% (両側検定なので半分) である。

{Type I Error, 0.0252179, Type II Error, 0.254809}



他方、帰無仮説が誤っているのに、それを採択してしまう誤りを**第 2 種の誤り**と呼ぶ。例えば、エマさんの検査値が 40 だったとする。そして、エマさんは本当は罹患していたとする。

両側検定、有意水準 $\alpha = 5\%$ として、以下のように帰無仮説を立ててみる。

帰無仮説：エマさんは健康である。

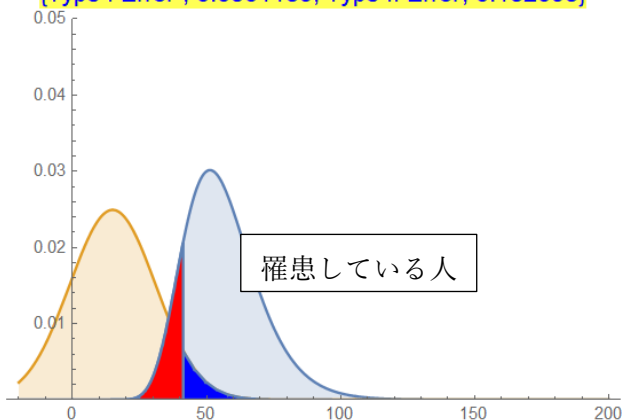
対立仮説：エマさんは罹患している。

よって、帰無仮説は棄却されない。しかし、エマさんは本当は罹患しているのに、第 2 種の誤りとなる。上の図で、 $\alpha = 2.5\%$ の限界値、約 46.3 のところで境界線を引くと、罹患していたのに帰無仮説を採択して健康としまう確率は、濃い赤の面積で表される。その確率 β は約 25.5% である。確率としてかなり大きな値である。

限界値は約 46.3 なので、40 はそれ以下の値である。

本当は罹患しているのに健康と誤診してしまっは大変なので、上記の濃い青の面積を大きくして、濃い赤の面積を小さくすることを考える。例えば、 α を 10% に増やそう。以下の図で濃い青の確率は 5% となっている。限界値は約 41.3 である。限界値を 46.3 から 41.3 に減じると、第 2 種の誤り (罹患しているのに、健康と判断) は約 13.2% に減った。

{Type I Error, 0.0501139, Type II Error, 0.132093}



人間ドックの結果を見て、「この位の数値で C 評価をつけるとはひどい」と感じることもあるが、病気を見逃すほうが深刻なので、厳格な境界値を設定するのであろう。

終わり

