Relating insomnia symptoms and genetic data
Research notes

Cüneyt
<cuneyt.guzey@ntnu.no>

Daniela
<daniela.bragantini@ntnu.no>

Luca
<piero.mana@ntnu.no>

Yasser
<yasser.roudi@ntnu.no>

Draft of 23 August 2018 (first drafted 22 August 2018)

Research notes

1 Preliminary remarks about Bayesian probability theory

Bayesian probability theory is not just a set of new, better recipes meant to replace old ones. It also requires a different – and simpler – mindset about problems of inference.

The only purpose of Bayesian theory is to give the probability of some statements – more exactly, 'propositions' (Copi et al. 2014; Barwise et al. 2003) – given other statements that may concern data, facts, hypotheses. For example, Bayesian theory can tell us that hypothesis A has probability x given some data D and initial information I, while hypothesis B has probability y given the same conditions:

$$\mathrm{P}(A \mid D, I) = x, \qquad \mathrm{P}(B \mid D, I) = y.$$

That's all there is to it. We can then use these probabilities as we like; in particular, we can use them within decision theory to choose courses of action (Raiffa et al. 2000; Pratt et al. 1996; Sox et al. 2013). But notions like 'statistical significance', 'acceptance level', 'confidence', and similar are foreign to Bayesian theory; or at best they're just secondary notions.

Another important characteristic of Bayesian theory is that it's an extension of formal logic, the truth calculus. In formal logic, to prove a theorem we need some axioms to start from. These may partly include experimental facts or data, but they always also include assumptions that are purely conjectural. It's impossible to avoid this conjectural element. This impossibility is well known in modern science; we can quote Poincaré [quote][ref], Einstein [quote][ref], Medawar [quote][ref], Jeffreys [quote][ref].

Likewise, in probability theory we need to specify initial probabilities. These may originate in data, but they always also include additional assumptions. The motto ‘let the data speak for themselves’ is simply impossible.

The difference between Bayesian methods and traditional methods is *not* that the former need additional assumptions while the latter don’t. Rather, Bayesian methods make these assumptions explicit, while traditional methods hide them. This is the reason why many traditional results can be obtained as special cases of Bayesian ones.

The previous remarks may appear pedantic, but they’re important lest we misuse Bayesian methods.

2 When genes keep you awake...

[Luca’s memoranda:]

- Use of partial exchangeability *has to* distinguish also between men and women: see Gehrman et al. (2013 p. 327).
- This study could also be used to detect most relevant genes, by eliminating them in turn (and in pairs etc) and checking the ensuing predictions.
- Is it computationally possible to use a ‘nonparametric model’? It would avoid unwarranted assumptions and phenomena like overtranining.

Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Barwise, J., Etchemendy, J. (2003): *Language, Proof and Logic*. (CSLI, Stanford). Written in collaboration with Gerard Allwein, Dave Barker-Plummer, Albert Liu. First publ. 1999.

Copi, I. M., Cohen, C., McMahon, K. (2014): *Introduction to Logic*, 14th ed. (Pearson, Harlow, UK). First publ. 1953.

Gehrman, P. R., Pfeiffenberger, C., Byrne, E. M. (2013): *The role of genes in the insomnia phenotype*. **8**3, 323–331.

Pratt, J. W., Raiffa, H., Schlaifer, R. (1996): *Introduction to Statistical Decision Theory*, 2nd pr. (MIT Press, Cambridge, USA). First publ. 1995.

Raiffa, H., Schlaifer, R. (2000): *Applied Statistical Decision Theory*, reprint. (Wiley, New York). First publ. 1961.

Sox, H. C., Higgins, M. C., Owens, D. K. (2013): *Medical Decision Making*, 2nd ed. (Wiley, New York). First publ. 1988.