Derivation of Kerridge's law

Andrew Fowlie

Nanjing Normal University

April 2021

The notation and presentation of ref. [1] make the proof and its implication hard for me to follow. Here I present a proof with missing steps filled in and without assuming equal prior plausibility of the models.

Let us suppose we have k hypotheses. The $k = 1$ hypothesis is true and the remaining $k - 1$ are false. They have prior probabilities π_i for $i = 1 \cdots k$. We can in fact reduce the scenario two two models: the true model with prior probability $\pi(T)$ and a mixture model of the false models, with prior probability $\pi(F) = \sum_{i=2}^{k} \pi_i$, so we proceed from that point.

First, consider the Bayes factor, B, in favour of the false model,

$$B = \frac{p\left(D \mid F\right)}{p\left(D \mid T\right)}. \tag{1}$$

Consider the constraint that the posterior of the true model is less than p, $p\left(T \mid D\right) \leq p$. By Bayes theorem alone, this implies

$$B \geq \left(\frac{1-p}{p}\right) \cdot \frac{\pi(T)}{\pi(F)} \tag{2}$$

Now consider a sum or integral over the region of sampling space in which

1. Sampling has stopped and
2. The posterior of the true model is less than p, $p\left(T \mid D\right) \leq p$.

We denote that sum or integral by $\sum^\star$. We perform that sum in

$$\sum^\star B p\left(D \mid T\right) \tag{3}$$

By Eq. 2 we must have

$$\sum^\star B p\left(D \mid T\right) \geq \left(\frac{1-p}{p}\right) \cdot \frac{\pi(T)}{\pi(F)} \cdot \sum^\star p\left(D \mid T\right) \tag{4}$$

By simply rewriting it using the definition of the Bayes factor in Eq. 1 though we must have

$$\sum^{\star} Bp\left(D \mid T\right) = \sum^{\star} p\left(D \mid F\right) \leq 1 \tag{5}$$

since we are summing over only part of the sampling space. Combing the Eqs. 4 and 5,

$$\left(\frac{1-p}{p}\right) \cdot \frac{\pi(T)}{\pi(F)} \cdot \sum^{\star} p\left(D \mid T\right) \leq \sum^{\star} Bp\left(D \mid T\right) \leq 1 \tag{6}$$

and so

$$\sum^{\star} p\left(D \mid T\right) \leq \left(\frac{1-p}{p}\right) \cdot \frac{\pi(F)}{\pi(T)} \tag{7}$$

If there are k equally plausibly hypothesis, $k-1$ of which are false, the prior odds factor would be $k-1$, recovering the result of Kerridge.

Finally, note again that $\sum^{\star}$ denotes a sum or integral over parts of the sampling space in which *i)* sampling has stopped and *ii)* $p\left(T \mid D\right) \leq p$. This means that we can write it in words as

$$P\left(\text{Sampling stopped and posterior probability of true hypothesis less than } p \mid \text{true hypothesis}\right)$$
$$\leq \left(\frac{1-p}{p}\right) \cdot \left(\text{Prior odds in favour of set of false hypotheses}\right) \tag{8}$$

In physics, we often don't consider optional stopping or stopping rules, in which case we can simply write

$$\begin{aligned} &P\left(\text{Posterior probability of true hypothesis less than } p \mid \text{true hypothesis}\right) \\ &\leq \left(\frac{1-p}{p}\right) \cdot \left(\text{Prior odds in favour of set of false hypotheses}\right) \end{aligned} \tag{9}$$

Kerridge's law gives a bound on the rate of misleading inferences in Bayesian model selection. Remarkably, the bound doesn't depend on the stopping rules. With p-values, you can sample until by (the law of the iterated logarithm) you reach an arbitrarily small p-value and stop.

Here, you cannot sample to a foregone conclusion. Your stopping rule could be that you stop only if the posterior probability of true hypothesis is less than p. But the probability that you ever stop would be bounded by Kerridge's law. i.e., by

$$\left(\frac{1-p}{p}\right) \cdot \left(\text{Prior odds in favour of set of false hypotheses}\right) \tag{10}$$

You could very well be sampling forever.

References

[1] D. Kerridge. Bounds for the frequency of misleading bayes inferences. *Ann. Math. Statist.*, 34(3):1109–1110, 09 1963.