Significance Testing

Possible outcomes of a significance test, where we think of H0 as the control hypothesis, so that a rejection of H0 is a "positive" result:

	Reality: H0 is not true	**Reality: H0 is true**
Test result: significant	True positive	False positive (Type I error)
Test result: non-significant	False negative (Type II error)	True negative

Or

	Reality: H0 is not true	**Reality: H0 is true**
Test result: significant	$(1-\beta)[1-P(H0)]$	$\alpha P(H0)$
Test result: non-significant	$\beta[1-P(H0)]$	$(1-\alpha)P(H0)$

Standard tests with $p < 0.05$ mean that if H0 is true then one would get a result this extreme less than 5% of the time.

This is usually interpreted as "the probability of such a result is less than 0.05, if H0 were true". Some argue that such an interpretation is incorrect. Allowing for this interpretation it says:

$$\frac{P(False\ positive)}{P(False\ positive)+P(True\ negative)}=\alpha<0.05.$$

It certainly does not tell us the probability that H0 is true, or equivalently, that H0 is not true, which is what we really care about.

Note that the **power** of a statistical test, which is the probability of correctly rejecting H0 when it is not true is:

$$\frac{P(True\ positive)}{P(True\ positive)+P(False\ negative)}=1-\beta.$$

For any statistical test, we can simulate data according to $H0$ (so $P(H0)=1$ in our simulation) and determine α from how often our test produces a "positive" result from such data. Similarly, we can simulate data according to an alternative hypothesis, not $H0$ (so $P(H0)=0$ in our simulation) and determine β from how often our test produces a "negative" result from such data.

We are ultimately interested in how much more likely is our alternative hypothesis than our control hypothesis given a "positive" test result. Clearly this is never simply the same as α, as even in the simplest case (only two alternatives with equal prior likelihood, so $P(H0=1)$) the posterior probability of $H0$ given a positive test is $\frac{\alpha}{1-\beta+\alpha}$ so depends on the power of the test.

To proceed in a direction that allows us to calculate what we want to know, *i.e.*, the posterior $P(H0|data)$, a Bayesian approach is necessary.