Hypothesis testing: Frequentists vs Bayesians

David McKay gives a nice comparison between Frequentist and Bayesian hypothesis testing:

" Two vaccinations A and B are tested on a group of volunteers to prevent the "microsoftus" disease. Vaccination B is only a placebo. Of the 40 subjects, 30 are assigned to treatment A and 10 to treatment B. After one year, one of group A and three of group B got microsoftus. Is treatment A better than B?"

FREQUENTIST:

The hypothesis H_1: 'A and B are different'

is transformed into the null-hypothesis

H_0: 'A and B are exactly equally effective'

Please note that what you really want to know, namely 'A is better than B' is not even mentioned. Frequentists will now compute how much the data deviates from the null-hypothesis.

We have four observables:

F_{A+} : # of people given treatment A that got ill

F_{A-} : # " " " " " that remained healthy

F_{B+} : # " " " " B that got ill

F_{B-} : # " " " " B that remained healthy

Then, a Frequentist will form the χ^2:

$$\chi^2 = \sum_{i=\{A+,A-,B+,B-\}} \frac{(F_i - \langle F_i \rangle)^2}{\langle F_i \rangle}$$

Actually, professionals use Yate's correction:

$$\chi^2 = \sum_i \frac{(|F_i - \langle F_i \rangle| - 0.5)^2}{\langle F_i \rangle}$$

As the Null-hypothesis claims that A and B are equally effective, the rates f_+ and f_- for contracting / not contracting the disease are then same:

$$\langle F_{A+} \rangle = f_+ N_A \qquad \langle F_{A-} \rangle = f_- N_A$$

$$\langle F_{B+} \rangle = f_+ N_B \qquad \langle F_{B-} \rangle = f_- N_B$$

For our Null-hypothesis, there is only one degree of freedom: we know N_A and N_B and e.g. $F_{B+} = N_B - F_{B-}$ and that $f_+ = 1 - f_-$.

So f_- can be regarded as the only degree of freedom. A table of χ^2's will then tell you, how "significant" a χ^2 is given one degree of freedom.* A usual threshold of significance is 5%.

Furthermore, a Frequentist will now estimate f_+ and f_- from the data:

$$\hat{f}_+ = \frac{F_{A+} + F_{B+}}{N_A + N_B} \quad ; \quad \hat{f}_- = \frac{F_{A-} + F_{B-}}{N_A + N_B}$$

For our example:

$$\hat{f}_+ = \frac{4}{40} = \frac{1}{10} \quad ; \quad \hat{f}_- = \frac{29 + 7}{40} = \frac{36}{40} = \frac{9}{10}$$

and for one degree of freedom: $\underline{\underline{\chi^2_{0.05} = 3.84}}$ (from a table).

So we expect:

$$\langle F_{A+} \rangle = \frac{1}{10} \cdot 30 = 3 \quad ; \quad \langle F_{A-} \rangle = \frac{9}{10} \cdot 30 = 27$$

$$\langle F_{B+} \rangle = \frac{1}{10} \cdot 10 = 1 \quad ; \quad \langle F_{B-} \rangle = \frac{9}{10} \cdot 10 = 9$$

And the $\underline{\underline{\chi^2 = 5.93}}$

and with Yate's correction:

$$\underline{\underline{\chi^2_{Yates} = 3.33}}$$

So with Yate's correction, we accept the Null hypothesis, without, we don't.

In a medical paper, the researches (using Yates correction) might now say that "There was no significant effect of the substance A".

Yet, what he really means is that "If you repeated our experiment very often, and the two treatments were equally effective, then in more than 5% of these experiments would you find the same rates f_+ and f_- that we saw in the actual data."

But that is not what we would like to know. All I personally care for is if treatment A is better than treatment B?

BAYESIAN:

We would form Bayes theorem for the probabilities P_{A+} and P_{B+} for contracting prob(microsoftus given either treatment:

$$\text{prob}(P_{A+}, P_{B+} \mid \{F_i\}) = \frac{\text{prob}(\{F_i\} \mid P_{A+}, P_{B+})\,\text{prob}(P_{A+}, P_{B+})}{\text{prob}(\{F_i\})}$$

As a prior, we use flat $\text{prob}(P_{A+}, P_{B+}) = 1$

The likelihood function is

$$\mathrm{prob}(\{F_i\} \mid p_{A+}, p_{B+}) = \binom{N_A}{F_{A+}} p_{A+}^{F_{A+}} p_{A-}^{F_{A-}} \binom{N_B}{F_{B+}} p_{B+}^{F_{B+}} p_{B-}^{F_{B-}}$$

$$= \binom{30}{1} p_{A+}^{1} \cdot p_{A-}^{29} \binom{10}{3} p_{B+}^{3} p_{B-}^{7}$$

We can now answer the question: "how probable is it that p_{A+} is smaller than p_{B+}"?

So we compute

$$\mathrm{prob}(p_{A+} < p_{B+} \mid \mathrm{Data})$$

which is

$$\int_{p_{A+}=0}^{1} \int_{p_{B+}=p_{A+}}^{1} dp_{A+}\, dp_{B+}\, \mathrm{prob}(p_{A+}, p_{B+} \mid \{F_i\})$$

divided by (i.e. normalized) by the full probability
$\int_0^1\int_0^1 dp_{A+} dp_{B+}\, \mathrm{prob}(p_{A+}, p_{B+} \mid \{F_i\})$.

Numerical evaluation of the integral yields

$$\mathrm{prob}(p_{A+} < p_{B+} \mid \{F_i\}) = 0.99$$

Hence, there is a 99% chance that treatment A is better than B. I bet that patients will be interested in this and not in the frequentist's analysis! Sadly, when you visit a Doctor, he or she might judge the treatment on the basis of the ill-perceived statement that "Treatment A is not significantly better than a placebo"?

You can also answer another interesting question, namely: „How likely is it that treatment A is ten times more effective than the placebo B":
For this, you need to integrate

$$\int_{P_{B+}=0}^{1} \int_{P_{A+}=0}^{10 \cdot P_{B+}} dP_{A+}\, dP_{B+}\, \mathrm{prob}(P_{A+}, P_{B+} \mid \{F_i\})$$