Reanalysis of Haffenden et al (2001) data

David J. C. MacKay
www.inference.phy.cam.ac.uk/mackay/

March 24, 2004

In Haffenden *et al.* (2001), the following data were reported in an experiment to test whether humans can smell a difference between ordinary benzaldehyde and deuterated benzaldehyde. In each experiment, the subject was given one reference smell, then presented with two other smells, one the same chemical as the reference, and the other different. The subject was then asked to identify which of the two was the same as the reference. Of 30 subjects, 23 got it right. What evidence does this give in favour of the hypothesis that humans can smell the difference? [Haffenden et al use sampling theory and report a P-value of 0.008.]

If we assume that some humans can always smell the difference and some always guess at random, what can be inferred about the fraction f of humans who can smell the difference? [Haffenden et al suggest $47 \pm 32\%$.]

In two experiments with other compounds, the number of correct identifications were 18/30 and 19/30 respectively, for which they quote P-values of 0.181 and 0.1. What are the posterior probabilities?

Haffenden et al report that presentation order affected their results (data follow below). How strong is the evidence for this assertion?

Answers using Bayesian analysis

(See (MacKay, 2003) for review of the Bayesian method.)

For the first part, the two hypotheses are ($\mathcal{H}_0$) that the probability that a subject gets the answer right in any one trial is 0.5, independently in all trials; and ($\mathcal{H}_1$) that there is a probability $p > 0.5$ that a human will get the answer right. [In terms of the fraction f who smell the difference reliably, $p = f + (1-f)/2 = 1/2 + f/2$.] Assuming $\mathcal{H}_1$ is true, I'd be happy to slap a uniform prior on f, though if pressed I would perhaps go for a mixture of beta distributions, since some smelling abilities are genetically determined, and many pairs of smells are reliably distinguished by everyone (so $f \simeq 1$ has quite a lot of prior probability).

The evidences contributed by the $r/N = 23/30$ result are

$$P(D|\mathcal{H}_0) = \binom{N}{r} 1/2^N \qquad (1)$$

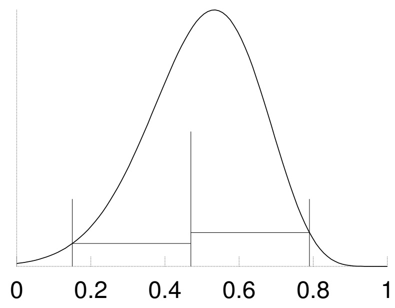

Figure 1: Posterior distribution of f, assuming $\mathcal{H}_1$ to be true.

$$P(D|\mathcal{H}_1) = \binom{N}{r} \int_{p=1/2}^{p=1} \mathrm{d}p\, 2\, p^r (1-p)^{N-r} \tag{2}$$

(assuming the uniform prior $P(p) = 2$ for $p \in (1/2, 1)$).

$$\frac{P(D|\mathcal{H}_0)}{P(D|\mathcal{H}_1)} = \frac{9 \times 10^{-10}}{1 \times 10^{-8}} = 0.059 \simeq 1/17. \tag{3}$$

Thus the data (23/30) give evidence about 34 to 1 in favour of $\mathcal{H}_1$.

The posterior distribution of f, assuming $\mathcal{H}_1$ to be true, is

$$P(f|D, \mathcal{H}_1) \propto (1/2 + f/2)^{23} (1/2 - f/2)^7 \tag{4}$$

This posterior distribution is plotted in figure 1 along with the sampling theory answer $47 \pm 32\%$, indicated by the thin straight lines.

The other datasets give the following evidence: (18/30) gives evidence 1.4 to 1 in favour of $\mathcal{H}_0$; (19/30) gives evidence 1.2 to 1 in favour of $\mathcal{H}_1$.

Presentation effect

Haffenden et al subdivided their data in accordance with presentation order and said “the ordering of the samples seemed to have influenced the results”. The data were

order	1	2	3	4
resulting number correct/trials	5/5	7/10	2/5	9/10

Let $\mathcal{H}_2$ be the hypothesis that assumes that there are four separate values of p, one for each condition, all greater than 1/2, and that assigns uniform priors for these four parameters; compare $\mathcal{H}_2$ with $\mathcal{H}_1$.

Solution:

with the data now being $D = \{r_t\} = (5, 7, 2, 9)$ given $\{N_t\} = (5, 10, 5, 10)$, the evidences are

$$P(D|\mathcal{H}_1) = \int_{p=1/2}^{p=1} \prod_t \binom{N_t}{r_t} \mathrm{d}p\, 2\, p_t^{r_t} (1-p_t)^{N_t - r_t} \tag{5}$$

$$P(D|\mathcal{H}_2) = \prod_t \binom{N_t}{r_t} \int_{p_t=1/2}^{p_t=1} \mathrm{d}p_t \, 2 \, p_t^{r_t} (1-p_t)^{N_t - r_t} \tag{6}$$

so

$$\frac{P(D|\mathcal{H}_1)}{P(D|\mathcal{H}_2)} = \frac{\int_{p=1/2}^{p=1} \mathrm{d}p \, 2 \, p^r (1-p)^{N-r}}{\prod_t \int_{p_t=1/2}^{p_t=1} \mathrm{d}p_t \, 2 \, p_t^{r_t} (1-p_t)^{N_t - r_t}} = \frac{1}{2.9}. \tag{7}$$

So the data give evidence of about 3 to 1 in favour of the hypothesis $\mathcal{H}_2$ that humans can distinguish the molecules, *and* there is an order-dependence, compared to the hypothesis that they can distinguish the samples and there is no order-dependence. (A not very strong result: plausible, but don't bet your house on it!)

Finally, if we wish to compare $\mathcal{H}_2$ to the null hypothesis, we do this not by "omitting affected sequences" but simply by comparing $P(D|\mathcal{H}_2)$ with $P(D|\mathcal{H}_0)$. The result is

$$\frac{P(D|\mathcal{H}_2)}{P(D|\mathcal{H}_0)} = \frac{98}{1}. \tag{8}$$

So the data are roughly 100 to 1 in favour of $\mathcal{H}_2$.

Appendix

More presentation order data, from table 8C.

Carbonyl-^{13}C				
order	1	2	3	4
resulting number correct/trials	5/8	7/7	5/8	2/7

Ring-$^{13}C_6$				
order	1	2	3	4
resulting number correct/trials	7/10	3/5	4/10	4/5

References

Haffenden, L. J. W., Yaylayan, V. A., and Fortin, J. (2001) Investigation of vibrational theory of olfaction with variously labelled benzaldehydes. *Food Chemistry* **73** (1): 67–72.

MacKay, D. J. C. (2003) *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press. Available from `http://www.inference.phy.cam.ac.uk/mackay/itila/`.