Appendix to the ISWC 2016 papers "Are Names Meaningful?"

Steven de Rooij, Wouter Beek, Peter Bloem, Frank van Harmelen, and Stefan Schlobach
{s.rooij,w.g.j.beek,p.bloem,frank.van.harmelen,k.s.schlobach}@vu.nl

Dept. of Computer Science, VU University Amsterdam, NL

Instead of reporting the usual p-value, we use the observed likelihood ratio λ itself which *bounds the p-value from above*[1]:

$$\begin{aligned} p = P_0(\Lambda \leq \lambda) &= P_0\left(\frac{P_1(Y_{1:n}|X_{1:n})}{P_0(Y_{1:n})} \geq \frac{1}{\lambda}\right) \\ &\leq \lambda \cdot E_{P_0}\left[\frac{P_1(Y_{1:n}|X_{1:n})}{P_0(Y_{1:n})}\right] = \lambda \end{aligned}$$

While p may be difficult, or impossible to compute, we can compute Λ easily, and report it as a upper bound on p.

In order to use this simple test we need to specify the distributions P_0 and P_1 exactly. The former should be the true distribution in case the null hypothesis is true and X and Y are independent. This distribution is unknown, but the probability of the data can be *overestimated* by its probability under the empirical distribution:

$$\hat{P}(Y = y) = \frac{|\{i \mid Y_i = y\}|}{n}.$$

The following inequality shows why this is an overestimation; here H denotes Shannon entropy and D the Kullback-Leibler divergence [1]. The inequality uses the nonnegativity of D:

$$\begin{aligned} \log P_0(Y_{1:n}) &= \sum_{i=1}^{n} \log P_0(Y_i) = n \sum_{y \in \mathcal{Y}} \hat{P}(y) \log P_0(Y_i) \\ &= -nH(\hat{P}) - nD(\hat{P} \| P_0) \\ &\leq -nH(\hat{P}) = \log \hat{P}(Y_{1:n}). \end{aligned}$$

Let $\hat{\Lambda}$ denote the likelihood ratio with respect to $\hat{P}$ instead of P_0 and let $\hat{\lambda}$ denote the observed value of $\hat{\Lambda}$. It is important that, even though our $\hat{\lambda}$ may be inaccurate, it satisfies $\hat{\lambda} \geq \lambda$ and is therefore more conservative when interpreted as a p-value than λ itself. In conclusion, we will report the value of the test statistic $\hat{\lambda}$, which is a conservative but valid p-value.

[1] In this equation we use Jensen's inequality [1] and the last equality is verified by expanding the expectation.

References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley-Interscience (2006)