Relative distance comparisons with confidence judgements
Supplementary material

Stefan Mojsilovic* Antti Ukkonen†

1 Proof of Theorem 4.1

We first derive the formula for the expected value of the estimator, and then show how it can be upper bounded by assuming the constraint on s_i given in the theorem. In the following, we shorten $d_{u_j v_j}$ as d_j.

The estimator for d_j is defined as:

$$d_j = \prod_{i \in S} 2(1-p_i) \prod_{i \in L} \frac{1}{2(1-p_i)}. \tag{1.1}$$

Since we have assumed the noise terms to be independent, we have

$$E[d_j] = \prod_{i \in S} E[2(1-p_i)] \prod_{i \in L} E[\frac{1}{2(1-p_i)}].$$

Also, we assume $p_i = p_i^* + \varepsilon_i$. Of these, only ε_i is a random variable, others are constants. Since $\varepsilon_i \sim \text{Unif}(-s, s)$, we have $E[\varepsilon_i] = 0$. Thus, by linearity of expectation we have $E[2(1-p_i)] = 2(1-E[p_i]) = 2(1-(p_i^* + E[\varepsilon_i])) = 2(1-p_i^*)$. For the second part, we have

$$\begin{aligned} E[\frac{1}{2(1-p_i)}] &= \int_{-s}^{s} 2(1-(p_i^*+\varepsilon_i))\frac{1}{2s} d\varepsilon_i \\ &= \frac{1}{4s}\log\left(\frac{1-p_i^*+s}{1-p_i^*-s}\right), \end{aligned}$$

which yields

$$E[d_j] = \prod_{i \in S} 2(1-p_i^*) \prod_{i \in L} \frac{1}{4s}\log\left(\frac{1-p_i^*+s}{1-p_i^*-s}\right).$$

Now consider the inequality $E[d_j] \le \alpha d_j^*$. Since the product $\prod_{i \in S} 2(1-p_i^*)$ (that corresponds to the short pairs) appears in both the expression for $E[d_j]$ and d_j^*, it can be divided out leaving us with

$$\prod_{i \in L} \frac{1}{4s}\log\left(\frac{1-p_i^*+s}{1-p_i^*-s}\right) \le \alpha \prod_{i \in L} \frac{1}{2(1-p_i^*)},$$

which can be rearranged as

$$\prod_{i \in L} \underbrace{\frac{1}{4s}\log\left(\frac{1-p_i^*+s}{1-p_i^*-s}\right) 2(1-p_i^*)}_{f(p_i^*)} \le \alpha.$$

We then derive an upper bound for $f(p_i^*)$ by using the bound $\log x \le x - 1$ to obtain

$$\frac{1}{4s}\log\left(\frac{1-p_i^*+s}{1-p_i^*-s}\right) \le \frac{1}{2(1-p_i^*-s)},$$

meaning

$$f(p_i^*) \le \frac{1-p_i^*}{1-p_i^*-s}.$$

This yields

$$\prod_{i \in L} f(p_i^*) \le \prod_{i \in L} \frac{1-p_i^*}{1-p_i^*-s} \le \prod_{i=1}^{j} \frac{1-p_i^*}{1-p_i^*-s} \le \alpha,$$

where the 2nd inequality follows from $\frac{1-p_i^*}{1-p_i^*-s} \ge 1$, and the fact that $|L| \le j$. Now, the rightmost inequality above holds when for all i

$$\frac{1-p_i^*}{1-p_i^*-s} \le \alpha^{1/j}.$$

Allowing s to depend on i, and solving the above for s_i gives

$$s_i \le -\left(1 - \frac{1}{\alpha^{1/j}}\right) p_i^* + \left(1 - \frac{1}{\alpha^{1/j}}\right),$$

which is the bound for s_i required in the theorem.

2 Proof of Theorem 4.2

Again, we shorten $d_{u_j v_j}$ as d_j. We use Chebyshev's inequality to determine a similar relationship between s and p_i as in Theorem 4.1, so that the probability of the estimates being "far" from their expected value (not the "true" d_j^*!) is bounded by q.

$$\begin{aligned} Pr(|d_j - E[d_j]| \ge aE[d_j]) &\le \frac{Var[d_j]}{a^2 E[d_j]^2} \\ &= \frac{1}{a^2}\left(\frac{E[d_j^2]}{E[d_j]^2} - 1\right) \le q. \end{aligned}$$

*University of Helsinki, Department of Computer Science
†University of Helsinki, Department of Computer Science

$$
\begin{aligned}
\frac{E[d_j^2]}{E[d_j]^2} &= \frac{\prod_{i\in S} E[4(1-p_i)^2]}{\prod_{i\in S} 4(1-p_i^*)^2} \frac{\prod_{i\in L} E[\frac{1}{4(1-p_i)^2}]}{\prod_{i\in L}(\frac{1}{4s}\log(\frac{1-p_i^*+s}{1-p_i^*-s}))^2} \\
&= \prod_{i\in S}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right)\prod_{i\in L}\frac{2s(\frac{1}{(1-p_i^*-s)}-\frac{1}{(1-p_i^*+s)})}{\log(\frac{1-p_i^*+s}{1-p_i^*-s})^2} \\
&\leq \prod_{i\in S}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right)\prod_{i\in L}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right) \\
&\leq \prod_{i=1}^{j}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right).
\end{aligned}
$$

(It can be shown that the first inequality above holds when we constrain s_i using the same bound as in Theorem 4.1 with any $\alpha \in [1, 23.75]$. We consider this to be a reasonable range of αs, as normally one would prefer the bias to be small. This means that if the amount of noise is sensible, the inequality holds.)

Substituting the above upper bound into Chebyshev RHS:

$$
\frac{1}{a^2}\left(\prod_{i=1}^{j}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right)-1\right) \leq q
$$

which is equivalent to

$$
\prod_{i=1}^{j}\left(1+\frac{s}{\sqrt{3}(1-p_i^*)}\right) \leq a^2 q + 1
$$

Again we allow $s = s_i$ to depend on p_i^*. The above holds when

$$
\left(1+\frac{s_i}{\sqrt{3}(1-p_i^*)}\right) \leq (a^2 q + 1)^{1/j} = \gamma,
$$

which holds when

$$
s_i \leq -\sqrt{3}(\gamma-1)p_i^* + \sqrt{3}(\gamma-1).
$$