

Hypergeometric

Let R be created by sampling n_1 points from X and n_2 points from \bar{X} . For the Forbes measure we then get:

$$\begin{aligned}
 |R| &= n_1 + n_2 \\
 FDR &= n_2 / (n_1 + n_2) \\
 |R \cap Q| &\sim HG(|X|, |X \cap Q|, n_1) + HG(|\bar{X}|, |\bar{X} \cap Q|, n_2) \\
 E\left(\frac{|R \cap Q|}{|R||Q|}\right) &= \frac{n_1 |X \cap Q| / |X| + n_2 |\bar{X} \cap Q| / |\bar{X}|}{|Q| (n_1 + n_2)} \\
 &= (1 - FDR) \frac{|X \cap Q|}{|X||Q|} + FDR \frac{|\bar{X} \cap Q|}{|\bar{X}||Q|}
 \end{aligned}$$

I.e: the expected Forbes measure for R and Q is a weighted linear combination of the Forbes similarity between X and Q and \bar{X} and Q . But independent of the size of R .

For Jaccard we get:

$$\begin{aligned}
 |R \cap Q| &\sim HG(|X|, |X \cap Q|, n_1) + HG(|\bar{X}|, |\bar{X} \cap Q|, n_2) \\
 E(|R \cap Q|) &= HG(|X|, |X \cap Q|, n_1) + HG(|\bar{X}|, |\bar{X} \cap Q|, n_2) \\
 \frac{|R \cap Q|}{|R \cup Q|} &= \frac{|Q| + |R| - |Q \cup R|}{|Q \cup R|} \\
 \frac{|R \cap Q|}{|R \cup Q|} &\sim \frac{|R \cap Q|}{|Q| + |R| - |R \cap Q|}
 \end{aligned}$$

Notes

$$\begin{aligned}
 f(x) &= (x)/(N - x) \\
 f'(x) &= ((N - x) + x)/(N - x)^2 \\
 f'(x) &= (N)/(N - x)^2
 \end{aligned}$$