Rechenweg Aufgabe 9.1 a)

Um die Handrechnung mit Hilfe von Janus nachrechnen zu können, verwenden wir für die Definition der Entropie einer Wahrscheinlichkeitsverteilung den logarithmus naturalis. Die Entropie einer distribution in Janus ist somit definiert als:

$$H(ds) = \sum_{i=1}^{n} -p_i \ln p_i$$

n ist hier die Anzahl der Gewichte in der distribution ds, p_i entspricht dem i-ten distribution weight.

Die Berechnung der gewichteten Entropie ist definiert als:

$$\bar{H}(ds) = P(ds) * H(ds)$$

In Janus ist es jetzt so, dass wir an Stelle der Wahrscheinlichkeiten $P(ds)$ die absoluten Häufigkeiten (engl. counts) der distributions erhalten. Gegeben eine Menge von distributions kann man eine Schätzung für $P(ds)$ erhalten, in dem man durch die Summe der absoluten Häufigkeiten teilt, d.h. darauf normiert.

In dem zu berechnenden Beispiel sind zwei distributions $ds1$ und $ds3$ gegeben. Die absolute Häufigkeit von $ds1$ ist 2 und von $ds3$ ist sie 3. Daraus ergibt sich:

$$P(ds1) = \frac{2}{5} = 0,4$$
$$P(ds1) = \frac{3}{5} = 0,6$$

Gegeben $ds1$ und $ds3$ berechnen sich jetzt die gewichteten Entropien zu:

$$\bar{H}(ds1) = P(ds1) * H(ds1) = 0,4 * (-0,3 * \ln 0,3 - 0,7 * \ln 0,7) = 0,2443457208$$
$$\bar{H}(ds3) = P(ds3) * H(ds3) = 0,6 * (-0,8 * \ln 0,8 - 0,2 * \ln 0,2) = 0,3002414542$$

Die Wahrscheinlichkeitsfunktion für $ds1 + ds3$ erhält man als Summe aus den einzelnen Funktionen gewichtet mit den absoluten Häufigkeiten:

$$ds1 + ds3 = \{0,6; 0,4\}$$
$$P(ds1 + ds3) = 1,0$$

Damit gilt:

$$\bar{H}(ds1+ds3) = P(ds1+ds3)*H(ds1+ds3) = 1,0*(-0,6*\ln 0,6-0,4*\ln 0,4) = 0,673011667$$

Die normierte Entropiedistanz erhält man jetzt als:

$$\Delta\bar{H}(ds1, ds3) = \bar{H}(ds1 + ds3) - (\bar{H}(ds1) + \bar{H}(ds3)) = 0,128424492$$