B Mutual Information Metric

B.1 Entropy

To introduce the mutual information metric, we present a review of entropy. Let us define Φ as a random variable (RV), $P(\Phi)$ as the probability distribution of Φ and $\rho(\phi)$ as the probability density of Φ. So the entropy of Φ, $H(\Phi)$ is :

$$H(\Phi) = -E\left[\log\left(P(\Phi)\right)\right] \tag{8}$$

where E is the expectation. Rewriting this equation yields the following:

$$H(\Phi) = -\sum_{\phi\in\Phi} \rho(\phi)\log_2 \rho(\phi) \tag{9}$$

This is entropy and it is a measure of randomness. The more random a variable is, the more the entropy. Joint entropy is a statistic that summarizes the degree of dependence of RV Φ on another RV Ψ :

$$H(\Phi,\Psi) = -E_{\phi}\left[E_{\psi}\left[\log\left(\rho(\Phi,\Psi)\right)\right]\right] \tag{10}$$

Again rewriting this equation yields:

$$H(\Phi,\Psi) = -\sum_{\phi\in\Phi}\sum_{\psi\in\Psi} \rho(\phi,\psi)\log_2 \rho(\phi,\psi) \tag{11}$$

The conditional entropy is a statistic that summarizes randomness of Ψ given knowledge of Φ :

$$H(\Phi|\Psi) = -E_{\phi}\left[E_{\psi}\left[\log\left(\rho(\Phi|\Psi)\right)\right]\right] \tag{12}$$

Two random variables are considered to be independent if:

$$H(\Phi,\Psi) = H(\Phi) + H(\Psi) \tag{13}$$

However if there is any dependency then:

$$H(\Phi,\Psi) < H(\Phi) + H(\Psi) \tag{14}$$

Working through some math we can rewrite the above equation:

$$H(\Psi|\Phi) = \sum_{\phi\in\Phi} \rho(\phi) H(\Psi|\Phi=\phi) \tag{15}$$

$$= -\sum_{\phi\in\Phi} \rho(\phi) \sum_{\psi\in\Psi} \rho(\psi|\phi)\log_2 \rho(\psi|\phi) \tag{16}$$

$$= -\sum_{\phi\in\Phi}\sum_{\psi\in\Psi} \rho(\phi,\psi)\log_2 \rho(\psi|\phi) \tag{17}$$

Now, given the chain rule;

$$H(\Phi,\Psi) = H(\Phi) + H(\Psi|\Phi) \tag{18}$$

We can now flip the variables and get:

$$H(\Psi, \Phi) = H(\Psi) + H(\Phi|\Psi) \tag{19}$$

hence the conditional entropy from Equation 12 becomes;

$$H(\Phi|\Psi) = H(\Psi, \Phi) - H(\Phi) \tag{20}$$

Now that we have defined entropy, joint entropy and conditional entropy we can introduce mutual information (MI).

B.2 Mutual Information

Mutual Information (MI), between two random variables Φ and Ψ is:

$$MI(\Phi, \Psi) = H(\Psi, \Phi) - H(\Psi|\Phi) \tag{21}$$

$$= H(\Phi) + H(\Psi) - H(\Psi, \Phi) \tag{22}$$

This is a measure of the reduction of the entropy of Ψ given Φ [17,18].

Maximizing the mutual information is documented by [17,18]. If the two distributions are related functionally (through a geometrical transformation $\mathbf{T}(\Phi)$), and we only have an estimate of this transformation, then the mutual information depends on

1. The first distribution, from image Φ.
2. The second distribution, from image Ψ.
3. The estimated transformation that maps one onto the other.

The exact transformation that maps the first image u (the fixed model) onto the second image v (the moving image) gives rise to the largest mutual information. Mutual information then becomes an optimization criterion, optimized w.r.t. T:
Let $\Phi = u(\Phi), \Psi = v(T(\Phi))$ where $u(\Phi)$ is a voxel of the reference volume, $v(\Phi)$ is a voxel of the test volume, and $\Psi = v(T(\Phi))$ is the test volume voxel associated with the reference volume voxel [18,17].

$$MI(T) = H(u(\Phi)) + H(v(T(\Phi))) - H(u(\Phi), v(T(\Phi))) \tag{23}$$

This method proceeds by a classic gradient-descent optimization technique, and tries to find the transformation that gives the largest mutual information by taking small steps in the direction of the derivative of the criterion [18,17].

$$\frac{d}{dT} MI(T) = \frac{d}{dT} H(v(T(\Phi))) - \frac{d}{dT} H(u(\Phi), v(T(\Phi))) \tag{24}$$