Energy Statistics

$\mathcal{E}$-Stats : function of distances between stat. observations. Should be 0 if null hypothesis is true.

d-dim random sample $X_1, \dots, X_n$

kernel $h: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} h(X_i, X_j) \qquad h(a,b) = h(b,a)$$

$$V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \qquad h = h(|a-b|)$$

Energy distance: $\mathcal{E}(X,Y) = 2\mathbb{E}|X-Y| - \mathbb{E}|X-X'| - \mathbb{E}|Y-Y'|$

X' is an iid copy of X
Y' is an iid copy of Y

Prop. $\mathcal{E}(X,Y) = \frac{1}{c_d} \int_{\mathbb{R}^d} \frac{|\hat{f}(t) - \hat{g}(t)|^2 dt}{|t|^{d+1}}$

$$c_d = \frac{\pi^{(d+1)/2}}{\Gamma\left(\frac{d+1}{2}\right)}$$

$\mathcal{E}(X,Y) = 0 \iff X, Y \sim p(x)$. (same dist.)

If X, Y take values in a metric space with distance δ: $\mathcal{E}(X,Y) = 2\mathbb{E}\,\delta(X,Y) - \mathbb{E}\,\delta(X,X') - \mathbb{E}\,\delta(Y,Y')$
In this case, the previous prop. do not necessarily holds.

Generalization:

$$\mathcal{E}^{(\alpha)}(X,Y) = 2\mathbb{E}|X-Y|^\alpha - \mathbb{E}|X-X'|^\alpha - \mathbb{E}|Y-Y'|^\alpha$$

$$= \frac{1}{c(d,\alpha)} \int_{\mathbb{R}^d} dt \, \frac{|\hat{f}(t) - \hat{g}(t)|^2}{|t|^{d+\alpha}}$$

k-sample hypothesis $H_0: F_1 = F_2 = \dots = F_k$.

$A = \{a_1, \dots, a_{n_1}\}$, $B = \{b_1, \dots, b_{n_2}\}$

$$g_\alpha(A,B) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} |a_i - b_m|^\alpha \qquad 0 < \alpha \leq 2.$$

Testing for equal distr.

$$\mathcal{E}(X,Y) = \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} |x_i - y_m| - \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} |x_i - x_j| - \frac{1}{n_2^2} \sum_{l=1}^{n_2} \sum_{m=1}^{n_2} |y_l - y_m|$$

$X = \{x_1, \dots, x_{n_1}\}$
$Y = \{y_1, \dots, y_{n_2}\}$

$$T = \frac{n_1 n_2}{n_1 + n_2} \mathcal{E}$$

$$T_\alpha = \frac{N}{2} g_\alpha(A,A) \qquad W_\alpha = \sum_{j=1}^{K} \frac{n_j}{2} g_\alpha(A_j, A_j)$$

$$N = \sum_{j=1}^{K} n_j$$

$$\mathcal{E}^{(\alpha)}_{n_j, n_k}(A_j, A_k) = 2 g_\alpha(A_j, A_k) - g_\alpha(A_j, A_j) - g_\alpha(A_k, A_k)$$

$$S_\alpha = \sum_{1 \leq j < k \leq K} \frac{n_j n_k}{2N} \mathcal{E}^{(\alpha)}_{n_j, n_k}$$

$$T_\alpha = S_\alpha + W_\alpha$$

③

$$\boxed{g = \frac{1}{n^2} \sum_i \sum_j |a_i - a_j|^2 = g(A, A)}$$

$$d^2(a_i, a_j) = (a_{i1} - a_{j1})^2 + \dots + (a_{id} - a_{jd})^2$$

$$= (a_i - a_j)^T (a_i - a_j) \quad \underline{\text{Euclidean}}$$

$$g = \frac{1}{n^2} \sum_{i,j=1}^{n} (a_i - a_j)^T (a_i - a_j)$$

$$= \frac{1}{n^2} \sum_{i,j=1}^{n} z_{ij}^T z_{ij}$$

$$z_{12}^T z_{12} + z_{13}^T z_{13} + \dots$$

$$\begin{pmatrix} z_{11}^T & z_{12}^T & z_{13}^T & \dots \end{pmatrix} \begin{pmatrix} z_{11} \\ z_{12} \\ z_{13} \\ \vdots \end{pmatrix} = \sum_{ij} z_{ij}^T z_{ij}$$

$$\boxed{g = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij}^T z_{ij} = \frac{1}{n^2} Z^T Z}$$

$$Z = \begin{pmatrix} z_{11} \\ z_{12} \\ z_{13} \\ \vdots \\ z_{1n} \\ z_{21} \\ z_{22} \\ z_{23} \\ \vdots \\ z_{2n} \\ \vdots \end{pmatrix} \in \mathbb{R}^{n^2}$$

stack all differences into Z

FROM symmetry we might reduce this to

$$Z = \begin{pmatrix} z_{12} \\ z_{13} \\ \vdots \\ z_{1n} \\ z_{23} \\ z_{24} \\ \vdots \\ z_{2n} \\ \vdots \end{pmatrix} = \begin{pmatrix} z_{ij} \end{pmatrix} \text{ with } i<j$$

$$(n-1) + (n-2) + (n-3) + \dots 1$$

$$n(n-1) - \sum_{i=1}^{n-1} i$$

$$n(n-1) - \frac{(n-1)n}{2} = \frac{n(n-1)}{2}$$

$$\boxed{g = \frac{n-1}{2n} Z^T Z}$$

$$|a_i - a_j|^\alpha = (d(a_i, a_j))^\alpha = (\langle a_i | a_j \rangle^{1/2})^\alpha$$

$$D_\alpha(a_i, a_j) \equiv |a_i - a_j|^\alpha$$

So suppose we form a distance matrix

$$D = [D_\alpha(a_i, a_j)]$$

containing all these distances. Then the sum of all entries of this matrix can be written as $\vec{1}^T D \vec{1}$ ($\vec{1} = (1, 1, \dots)^T$). Thus

$$\boxed{g = \frac{1}{n^2} \vec{1}^T D \vec{1}} \qquad n = \vec{1} \cdot \vec{1}$$

~~Considering again~~ $\cancel{g = \frac{1}{n^2} \sum_i^n \sum_j |a_i - a_j|^2}$

~~Another way~~

consider again
$$g = \frac{1}{n^2} \sum_{i,j} |a_i - a_j|^2$$
$$= \frac{1}{n^2} \sum_{i,j} d(a_i, a_j)^2$$
$$= \frac{1}{n^2} \sum_{i,j} d_{ij}^2$$

If we form the matrix $D = (d_{ij})$ then we can also write

$$\boxed{g = \frac{1}{n^2} \mathrm{Tr}(D^T D)}$$

In the general case we can also do the following:

$g = \frac{1}{n^2} \mathbb{1}^T D \vec{\mathbb{1}}$. Insted form the Matrix $E = \vec{\mathbb{1}} \cdot \vec{\mathbb{1}}^T$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix}_{n\times 1} (1\ 1\ 1\ \cdots)_{1\times n}$$

$$= \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}_{n\times n}$$

$$ED = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}_{n\times n} \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & d_{nn} \end{pmatrix}_{n\times n}$$

$$= \begin{pmatrix} \underline{d_{11} + d_{21} + \cdots + d_{n1}} & d_{12} + d_{22} + \cdots d_{n2} \\ d_{11} + d_{21} + \cdots + d_{n1} & \underline{d_{12} + d_{22} + \cdots d_{n2}} \\ & & \ddots \end{pmatrix}$$

$$\therefore \boxed{g = \frac{1}{n^2} \operatorname{Tr}(ED)}$$

Moreover D is symmetric, thus ca be diagonalized

$Q^{-1} D Q = Q^T D Q = \Lambda \therefore \boxed{D = Q \Lambda Q^T}$

orthogonal $\boxed{Q Q^T = I}$

$$\mathbb{1}^T D \mathbb{1} = \mathbb{1}^T Q \Lambda Q^T \mathbb{1}$$
$$= (Q^T \cdot \vec{\mathbb{1}})^T \Lambda (Q^T \mathbb{1})$$

Quadratic Form.

$\cancel{= x^T A x}$

$$= \sum_{i=1}^{n} d_i q_i q_i^T$$

$$g = \sum_{i=1}^{n} \sum_{j=1}^{m} |a_i - b_j|^\alpha$$

$$D = \left(|a_i - b_j|^\alpha \right)_{n \times m}$$ Non quadratic Distance Matrix

we can express this as

$$g = \vec{1}_m D \vec{1}_m$$

or also as

$$\boxed{g = Tr(\vec{1}_m \vec{1}_n^T D) = Tr(E D)}$$