GSEA KS and KL Enrichment Statistics

Cumulative distributions for KS and KL GSEA (P, Q, U and R)

$$P_i^{g_k} = \frac{\sum_{h=1}^{i} |S_h|^\alpha w_h I_h}{\sum_{h=1}^{N} |S_h|^\alpha w_h I_h} \qquad Q_i^{g_k} = \frac{\sum_{h=1}^{i} (1 - I_h)}{(N - n_k)} \qquad U_i^{g_k} = \frac{\sum_{h=1}^{i} |S_h|^\alpha w_h (1 - I_h)}{\sum_{h=1}^{N} |S_h|^\alpha w_h (1 - I_h)} \qquad R_i^{g_k} = \frac{\sum_{h=1}^{i} |S_h|^\alpha w_h}{\sum_{h=1}^{N} |S_h|^\alpha w_h}$$

Standard (KS) GSEA:

- **KS_SUP**: Kolmogorov-Smirnov Supremum
- **KS_AUC**: Kolmogorov-Smirnov Area Under the Curve

Information Theoretical (KL) GSEA:

- **KL_PR**: Kullback-Leibler Divergence of P and R (KL-Li)
- **KL_PQ**: Kullback-Leibler Divergence of P and Q
- **KL_PRURm**: Kullback-Leibler Divergence of P and R minus Divergence of U and R (KL-Lio-)
- **KL_PRURp**: Kullback-Leibler Divergence of P and R plus Divergence of U and R (KL-Lio+)

Standard (KS) GSEA: KS_SUP

KS_SUP: Kolmogorov-Smirnov Supremum

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k}) = P_i^{g_k} - Q_i^{g_k}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \sup_{i=1,\dots,N} \Delta_i^{g_k}$$

Standard (KS) GSEA: KS_AUC

KS_AUC: Kolmogorov-Smirnov Area under the Curve

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k}) = P_i^{g_k} - Q_i^{g_k}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \frac{1}{N}\sum_{i=1}^{N}\Delta_i^{g_k}$$

Information Theoretical (KL) GSEA: KL_PR

KL_PR: Kullback-Leibler Divergence of P and R (KL-Li)

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, R_i^{g_k}) = D_{KL}(P_i^{g_k} \parallel R_i^{g_k}) = P_i^{g_k} \log \frac{P_i^{g_k}}{R_i^{g_k}}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, R_i^{g_k}) - \Phi_i^{g_k}(1 - P_i^{g_k}, 1 - R_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i^{g_k}$$

Information Theoretical (KL) GSEA: KL_PQ

KL_PQ: Kullback-Leibler Divergence of P and Q

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k}) = D_{KL}(P_i^{g_k} \parallel Q_i^{g_k}) = P_i^{g_k} \log \frac{P_i^{g_k}}{Q_i^{g_k}}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, Q_i^{g_k}) - \Phi_i^{g_k}(1 - P_i^{g_k}, 1 - Q_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i^{g_k}$$

Information Theoretical (KL) GSEA: KL_PRURm

KL_PRURm: Kullback-Leibler Divergence of P and R minus Divergence of U and R (KL-Lio-)

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, U_i^{g_k}, R_i^{g_k}) = D_{KL}(P_i^{g_k} \parallel R_i^{g_k}) - D_{KL}(U_i^{g_k} \parallel R_i^{g_k}) = P_i^{g_k} \log \frac{P_i^{g_k}}{R_i^{g_k}} - U_i^{g_k} \log \frac{U_i^{g_k}}{R_i^{g_k}}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, U_i^{g_k}, R_i^{g_k}) - \Phi_i^{g_k}(1 - P_i^{g_k}, 1 - U_i^{g_k}, 1 - R_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i^{g_k}$$

Information Theoretical (KL) GSEA: KL_PRURp

KL_PRURp: Kullback-Leibler Divergence of P and R plus Divergence of U and R (KL-Lio-)

Enrichment Statistic

$$\Phi_i^{g_k}(P_i^{g_k}, U_i^{g_k}, R_i^{g_k}) = D_{KL}(P_i^{g_k} \parallel R_i^{g_k}) + D_{KL}(U_i^{g_k} \parallel R_i^{g_k}) = P_i^{g_k} \log \frac{P_i^{g_k}}{R_i^{g_k}} + U_i^{g_k} \log \frac{U_i^{g_k}}{R_i^{g_k}}$$

Running Enrichment

$$\Delta_i^{g_k} = \Phi_i^{g_k}(P_i^{g_k}, U_i^{g_k}, R_i^{g_k}) - \Phi_i^{g_k}(1 - P_i^{g_k}, 1 - U_i^{g_k}, 1 - R_i^{g_k})$$

Enrichment Score

$$E_{g_k} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i^{g_k}$$