

Robertson-Sparck Jones Model

(Robertson & Sparck Jones 76)

$$\log O(R=1 | Q, D) \approx \sum_{i=1, d_i=q_i=1}^{Rank, k} \log \frac{p_i(1-q_i)}{q_i(1-p_i)} \quad (\text{RSJ model})$$

Two parameters for each term A_i :

$p_i = P(A_i=1 | Q, R=1)$: prob. that term A_i occurs in a relevant doc

$q_i = P(A_i=1 | Q, R=0)$: prob. that term A_i occurs in a non-relevant doc

How to estimate parameters?

Suppose we have relevance judgments,

$$\hat{p}_i = \frac{\#(\text{rel. doc with } A_i) + 0.5}{\#(\text{rel.doc}) + 1} \quad \hat{q}_i = \frac{\#(\text{nonrel. doc with } A_i) + 0.5}{\#(\text{nonrel.doc}) + 1}$$

“+0.5” and “+1” can be justified by Bayesian estimation