

RSJ Model: No Relevance Info

(Croft & Harper 79)

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

How to estimate parameters?

Suppose we do not have relevance judgments,

- We will assume p_i to be a constant
- Estimate q_i by assuming **all** documents to be **non-relevant**

$$\log O(R = 1 | Q, D) \stackrel{\text{Rank}}{\approx} \sum_{i=1, d_i=q_i=1}^k \log \frac{N - n_i + 0.5}{n_i + 0.5} \quad IDF' = \log \frac{N - n_i}{n_i}$$

N: # documents in collection

n_i : # documents in which term A_i occurs