

# Query Generation (→ Language Models for IR)

$$\begin{aligned} O(R=1|Q,D) &\propto \frac{P(Q,D|R=1)}{P(Q,D|R=0)} \\ &= \frac{P(Q|D,R=1)P(D|R=1)}{P(Q|D,R=0)P(D|R=0)} \\ &\propto \frac{P(Q|D,R=1)}{P(D|R=0)} \quad (\text{Assume } P(Q|D,R=0) \approx P(Q|R=0)) \end{aligned}$$

Query likelihood  $p(Q|D,R=1)$

Document prior

Assuming uniform prior, we have  $O(R=1|Q,D) \propto P(Q|D,R=1)$

Now, the question is how to compute  $P(Q|D,R=1)$  ?

Generally involves two steps:

- (1) estimate a language model based on D
- (2) compute the query likelihood according to the estimated model

**$P(Q|D, R=1)$  Prob. that a user who likes D would pose query Q. How to estimate it?**