1. see above
2. application of smoothing
    a. results:
        a 0.048
        the 0.24533333333333335
        from 0.034666666666666665
        retrieval 0.005333333333333333
        sun 0.08
        rises 0.07733333333333332
        in 0.17600000000000002
        BM25 0.0026666666666666666
        east 0.07200000000000001
        sets 0.07733333333333332
        west 0.07200000000000001
        and 0.10933333333333334

    b. u = .01
        a 0.00016348773841961853
        the 0.27263396911898274
        from 0.00011807447774750229
        retrieval 1.8165304268846504e-05
        sun 0.09087193460490463
        rises 0.0908628519527702
        in 0.1817983651226158
        BM25 9.082652134423252e-06
        east 0.09084468664850136
        sets 0.0908628519527702
        west 0.09084468664850136
        and 0.0909718437783833

        u = 100

        a 0.16216216216216217
        the 0.18018018018018017
        from 0.11711711711711711
        retrieval 0.018018018018018018
        sun 0.05405405405405406
        rises 0.04504504504504504
        in 0.16216216216216217
        BM25 0.009009009009009009
        east 0.02702702702702703
        sets 0.04504504504504504

west 0.02702702702702703
and 0.15315315315315314

As u increases weight increases. This makes sense because looking at the formula for Dirichlet prior smoothing u is multiplied by p(w|C), however it is not multiplied by anything in the denominator. Thus no matter what u is, as it increases the numerator will increase at a faster rate than the denominator, and thus we would expect p(w|d) to increase.

c. lambda = .1
a 0.0018
the 0.27169999999999994
from 0.0013000000000000002
retrieval 0.0002
sun 0.0905
rises 0.0904
in 0.18159999999999998
BM25 0.0001
east 0.0902
sets 0.0904
west 0.0902
and 0.0916
lambda =.5
a 0.09
the 0.22136363636363637
from 0.065
retrieval 0.01
sun 0.07045454545454546
rises 0.06545454545454546
in 0.1709090909090909
BM25 0.005
east 0.05545454545454546
sets 0.06545454545454546
west 0.05545454545454546
and 0.12545454545454546

lambda = .9
a 0.162
the 0.18027272727272728
from 0.117
retrieval 0.018000000000000002
sun 0.05409090909090909

rises 0.04509090909090909
in 0.1621818181818182
BM25 0.009000000000000001
east 0.02709090909090909
sets 0.04509090909090909
west 0.02709090909090909
and 0.1530909090909091
The Dirichlet prior has less smoothing than the Jelinek mercer, which makes sense because for Dirichlet prior the amount of smoothing is weighted by document length, and in this case the length of our document is quite short.

19. $$p(w|d) = \frac{c(w,d) + N \cdot p(w|C)}{|d| + N}$$

$$= \frac{|d|}{|d|+N} \frac{c(w,d)}{|d|} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= \lim_{d \to \infty} \frac{c(w,d)}{|d|+N} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= \lim_{d \to \infty} \frac{c(w,d)}{|d|+N} + 0$$

$$\approx \boxed{\frac{c(w,d)}{|d|}} \quad \blacksquare$$

$$\lim_{N \to \infty} \frac{|d|}{|d|+N} \frac{c(w,d)}{|d|} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= 0 + \lim_{N \to \infty} \frac{N}{N+|d|} \cdot p(w|C)$$

$$= p(w|C) \quad \blacksquare$$

19. $p(w|d) = \dfrac{c(w,d) + N \cdot p(w|C)}{|d| + N}$

$$= \frac{|d|}{|d|+N} \frac{c(w,d)}{|d|} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= \lim_{d \to \infty} \frac{c(w,d)}{|d|+N} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= \lim_{d \to \infty} \frac{c(w,d)}{|d|+N} + 0$$

$$\approx \boxed{\frac{c(w,d)}{|d|}} \blacksquare$$

$$\lim_{N \to \infty} \frac{|d|}{|d|+N} \frac{c(w,d)}{|d|} + \frac{N}{|d|+N} \cdot p(w|C)$$

$$= 0 + \lim_{N \to \infty} \frac{N}{N+|d|} \cdot p(w|C)$$

$$= p(w|C)$$

$\blacksquare$

39.

$$O(R=1 \mid Q, D) = \frac{P(R=1 \mid Q, D)}{P(R=0 \mid Q, D)} = \frac{P(Q, D \mid R=1)\, P(R=1)}{P(Q, D \mid R=0)\, P(R=0)}$$

Constant

$$\propto \frac{P(Q, D \mid R=1)}{P(Q, D \mid R=0)}$$

$$P(Q, D \mid R) = P(D \mid Q, R)\, P(Q \mid R)$$

$$\frac{P(D \mid Q, R=1)\, P(Q \mid R=1)}{P(D \mid Q, R=0)\, P(Q \mid R=0)}$$

Constant

Let $D = d_1, \ldots, d_k$, be 0 or 1 for their presence and absence in the document.

$$\prod_{i=1, d=1}^{k} \frac{P(A_i=1 \mid Q, R=1)}{P(A_i=1 \mid Q, R=0)} \cancel{\prod_{i=1, d=0}^{k} \frac{P(A_i=0 \mid Q, R=1)}{P(A_i=0 \mid Q, R=0)}}$$

absence of terms does not depend on relevance

$$\prod_{i=1, d_i=1}^{K} \frac{p(A_i=1|Q,R=1)}{p(A_i=1|Q,R=0)}$$

$$\text{Score}(Q,D) = \prod_{i=1, d_i=1}^{|V|} \left(\frac{P(w_i|Q,R=1)}{P(w_i|Q,R=0)}\right)^{c(w,D)}$$

$$\text{Score}(Q,D) = \sum_{i=1, d_i=1}^{|V|} c(w_i,D) \log\left(\frac{p(w_i|Q,R=1)}{p(w_i|Q,R=0)}\right)$$

we will need to estimate probability of relevance and non relevance

3b. $$p(w|Q,R=0) = \frac{\sum_{i=1}^{N} c(w,d)}{\sum_{i=1}^{N} |D_i|}$$

3c. $$\frac{c(w,Q)}{|Q|}$$

3d. $p(w \mid q, R=1) = (1-\lambda)\frac{c(w \mid Q)}{|Q|} + \lambda p(w \mid C)$

3E. $$\sum_{w \in V} \left(\log \left[\frac{(1-\lambda)\frac{c(w \mid Q)}{|Q|} + \lambda p(w \mid C)}{\frac{\sum_{i=1}^{N} c(w, d_i)}{\sum_{i=1}^{N} |D|}} \right] \right)$$

Yes it captures TF, IDF and document length normilization. IDF is in the denomination, as well as document length normilization.