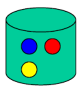
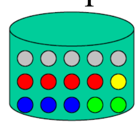
interpolation methods

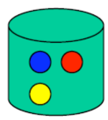
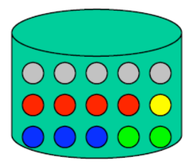
- Problem with all discounting methods:
  - discounting treats unseen words equally (add or subtract ε)
  - some words are more frequent than others
- Idea: use background probabilities
  - "interpolate" ML estimates with General English expectations (computed as relative frequency of a word in a large collection)
  - reflects expected frequency of events

ML estimate



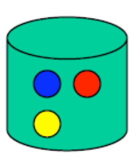
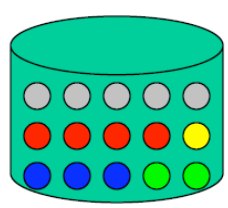
background probability



final estimate = λ  + (1–λ) 

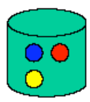
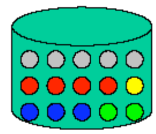
Jelinek Mercer smoothing

- Correctly setting **λ** is very important
- Start simple
  - set **λ** to be a constant, independent of document, query
- Tune to optimize retrieval performance
  - optimal value of **λ** varies with different databases, query sets, etc.

λ  + (1–λ) 

Dirichlet smoothing

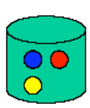
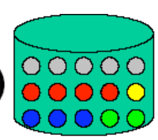
- Problem with Jelinek-Mercer:
  - longer documents provide better estimates
  - could get by with less smoothing
- Make smoothing depend on sample size
- N is length of sample = document length
- μ is a constant

$\underbrace{N / (N + \mu)}_{\lambda}$  + $\underbrace{\mu / (N + \mu)}_{(1-\lambda)}$ 

38

Witten-Bell smoothing

- A step further:
  - condition smoothing on “redundancy” of the example
  - long, redundant example requires little smoothing
  - short, sparse example requires a lot of smoothing
- Derived by considering the proportion of new events as we walk through example
  - N is total number of events = document length
  - V is number of unique events = number of unique terms in doc

$\underbrace{N / (N + V)}_{\lambda}$  + $\underbrace{V / (N + V)}_{(1-\lambda)}$ 

39