

# Developing the Quantum PRP

Extended abstract summarising  
“The QPRP for IR”, ICTIR 2009, and  
“Using the QPRP to Rank Interdependent Documents”, ECIR 2010

**Guido Zuccon**, Leif Azzopardi,  
Information Retrieval Group,  
University of Glasgow, UK



## The Probability Ranking Principle

- Rank documents in decreasing order of  $P(d)$ 
  - Proposed by Cooper [Cooper1971], formalised by Robertson [Robertson1977]
  - Optimal [Gordon&Lenk1999] if the assumptions are satisfied, e.g. ad-hoc retrieval
- Assumptions of the PRP
  - ...
  - Independent document relevance
  - Certainty in the estimation of probabilities
  - ...

## Critics to PRP

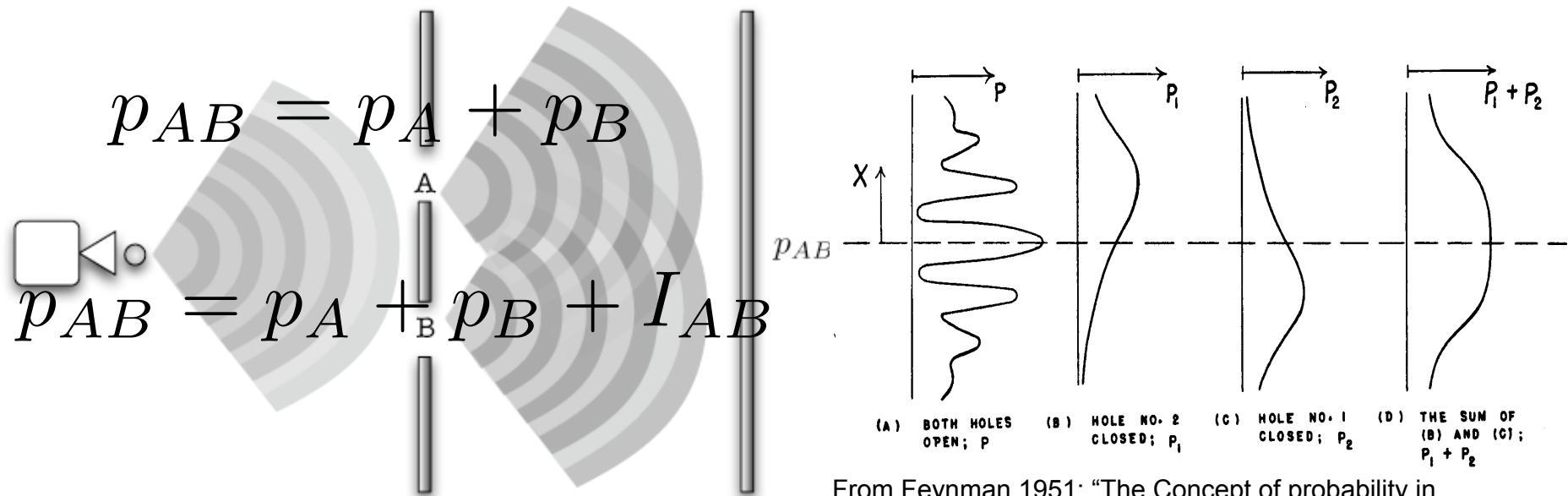
- But, is this always true?
  - Can we always safely assume that documents' relevance does not depend by previous seen documents?
    - Goffman 1964: On relevance as a measure
    - Chen&Karger 2006: Less is more
    - Gordon&Lenk 1999: When is the PRP suboptimal?

## Critics to PRP

- Several attempts to overcome PRP's limitations in particular scenarios
  - Interdependent document relevance: MMR [Carbonell&Goldstein1998, Zhai2003], iPRP [Fuhr2008]
  - Uncertainty in  $P(d)$  estimations: PT[Wang&Zhu2009]
- ...and to evaluate approaches&strategies for interdependent document relevance
  - Subtopics retrieval [Zhai2003], faceted topics retrieval [Carterette&Chandar2009]
  - novelty&diversity measures [Clarke&al.2008,2009]
  - Test collections (ClueWeb, imageCLEF)

## Towards Quantum PRP

- Based on analogy with double slits experiments



From Feynman 1951: "The Concept of probability in Quantum Mechanics"

- In the analogy, quantum events are about  $\mathbb{R}$  in the ranking
- "passing through slit" -> event
- Interfering documents: interference occurs at the events interference relevance level

Further readings: using events in [Kjersbergen 2004, Wildows 2004, Bruza 2004, 2005, 2009]

## The Quantum PRP

- Ranking rule:

$$d_i = \arg \max \left( P(d) + \sum_{d' \in RA} I_{d,d'} \right)$$

- Accounts for interdependent document relevance through interference

## The interfere term

in the range

$$\left[ -\sqrt{P(d)P(d')}, \sqrt{P(d)P(d')} \right]$$

Depends from cosine of phase-difference of complex numbers

$I_{d,d'}$

$$I_{d,d'} = \sqrt{P(d)P(d')} \cos \theta_{d,d'}$$

Needs to be approximated

## Empirical Investigation

- Based on TREC 6, 7, 8 interactive/subtopics collection

| Models      | <i>S-r@5</i>               | <b>S-r@10</b>             | <b>S-r@20</b>             | <b>S-r@50</b>             | <b>S-MRR@100</b>           | <b>S-p@.1</b>              | <b>S-p@.2</b>              | <b>S-p@.5</b>              |
|-------------|----------------------------|---------------------------|---------------------------|---------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| <b>PRP</b>  | 0.2466                     | 0.3900                    | 0.4962                    | 0.6034                    | 0.0086                     | 0.3968                     | 0.3062                     | 0.1941                     |
| <b>MMR</b>  | 0.2697<br>(+8.56%)         | 0.3540<br>(-10.19%)       | 0.4795<br>(-3.49%)        | 0.6032<br>(-0.02%)        | 0.097<br>(+11.39%)         | 0.4203<br>(+5.59%)         | 0.2876<br>(-6.48%)         | 0.1964<br>(+1.17%)         |
| <b>PT</b>   | 0.2791<br>(+11.63%)        | 0.3654<br>(-6.75%)        | 0.4444<br>(-11.66%)       | 0.5494<br>(-9.82%)        | 0.0130<br>(+33.47%)        | <b>0.4587</b><br>(+13.48%) | 0.2915<br>(-5.05%)         | 0.1769<br>(-9.73%)         |
| <b>QPRP</b> | <b>0.3093</b><br>(+20.25%) | <b>0.4063</b><br>(+3.99%) | <b>0.5026</b><br>(+1.27%) | <b>0.6186</b><br>(+2.45%) | <b>0.0177</b><br>(+51.31%) | 0.4237<br>(+6.33%)         | <b>0.3446</b><br>(+11.14%) | <b>0.2362</b><br>(+17.81%) |

Table 1: Subtopic retrieval performance when  $n = 200$ : where PT ( $b = 4$ ,  $\delta^2 = 10^{-5}$ ) and MMR ( $c = 0.5$ ) are optimized for S-r@5 measure.

- Accounting for dependencies improves subtopics retrieval performances
- PRP is sub-optimal in this case
- The QPRP improves the best

More results in [Zuccon&Azzopardi,ECIR2009] and <http://www.dcs.gla.ac.uk/~guido/qprresults.html>

## Questions?

Thanks to :

- Keith van Rijsbergen for his collaboration, support and mentoring
- and to Massimo Melucci, Giorgio Di Nunzio and Emanuele Di Buccio for the useful comments and discussions



Guido Zuccon  
Information Retrieval Group  
University of Glasgow  
guido@dcs.gla.ac.uk