OpenSource Connections

An approach to modelling implicit user feedback for optimizing e-commerce search

opensourceconnections.com

René Kriegler
Haystack - The Search Relevance Conference
Charlottesville, 27 April 2022

Who am I?

- Director E-commerce Search at OpenSource Connections
- Worked in search for 15 years
- Focus on e-commerce search, worked with some of Germany's top 10 online retailers
- Co-Founder/-Organiser of MICES - Mix-Camp Ecommerce Search (https://mices.co)
- Lucene, Solr, Elasticsearch
- Maintainer of Querqy, co-initiator of Chorus

Buying, Having, Being - Aspects of search quality in e-commerce

The What Having

The right type of item, the utility of the item

Modelling in search

Core retrieval algorithm matching the semantics of the query (‘smartphone 64 gb’)

Evaluation

Manual evaluation

The Purchase Buying

Readiness to spend money on that item, making the purchase decision for a concrete offer

Modelling in search

Mostly explicit, numerical factors that influence the buying decision (price, delivery time, reviews, ...)

Evaluation

Implicit in user behaviour

The Implied Being

(Social) Implications on the self

Modelling in search

Implicit/latent decision factors (consumer buys locally, prefers brand X, colour green, open to try out things, ...)

Evaluation

Implicit in user behaviour

‘Buying, Having, and Being’ is taken from the subtitle of Michael R. Solomon. *Consumer Behavior. Buying, Having, and Being*. 2006.

Judgment modelling

Problem:

- How much will a consumer be satisfied by a product/offer as a result to a given search query? (Non-ecommerce search: How relevant is a given document for a given query?)

Approach:

- Derive search quality judgments from observed user behaviour

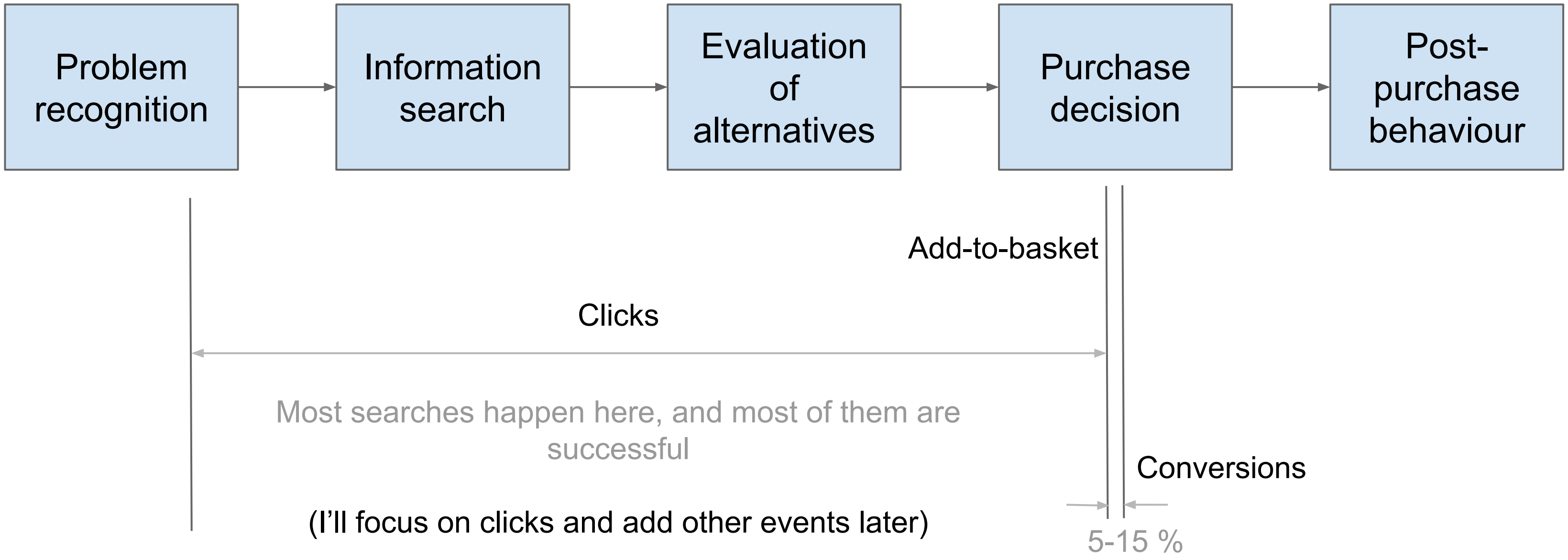
Aspects of judgment modelling

Model shall reflect these aspects:

- Event probability: a probability that tells us that how likely it is that a given product/offer/document is a good answer to the query (click rate, conversion rate)
- Certainty: How sure can we be about the event probability? (few observations, high variance -> uncertainty)
- Context: try to eliminate influence of position, device, grid size, ...

Event types

Philip Kotler, Kevin Lane
Marketing Management
1997
Peter Morville
Ambient Findability, 2005



(I’ll focus on clicks and add other events later)

Let's get started: CTR as probability

$$P(C = 1) = \rho$$

Inspired by
Chuklin et al.
Click Models for Web Search
2015

$$P(C_{qd} = 1) = ctr_{qd} = \frac{c_{qd}}{v_{qd}}$$

C: 1 if clicked, 0 if not

c_{qd}: observed clicks for query-document pair qd

v_{qd}: tracked views for query-document pair qd

CTR as probability

clicks	views	ctr
20	100	0.20
0	100	0.00
0	1	0.00
1	1	1.00
99	100	0.99

} ?

} ?

} ?

CTR as probability

clicks	views	ctr
20	100	0.20
0	100	0.00
0	1	0.00
1	1	1.00
99	100	0.99

} ?

Uncertainty! We would expect the ctr to change with the next event that we collect!

Probability: working with an expected value

Model an <u>expected probability</u> as a function of parameters *alpha* and *beta*...

$$E[ctr] = \overline{ctr} = \frac{1}{N}\sum_{n=1}^{N} ctr_{qd_n}$$

$$E[ctr] = \frac{\alpha}{\alpha + \beta}$$

Probability: working with an expected value

Model an expected probability as a function of parameters *alpha* and *beta*...

$$E[ctr] = \frac{\alpha}{\alpha + \beta}$$

... and update this with the observed clicks and views

$$y = \frac{\alpha + c_{qd}}{\alpha + \beta + v_{qd}}$$

Probability: working with an expected value

E[ctr]=0.1 | alpha=1, beta=9

clicks	views	ctr	y
20	100	0.20	0.19
0	100	0.00	0.01
0	1	0.00	0.09
1	1	1.00	0.18
99	100	0.99	0.91

$$y = \frac{\alpha + c_{qd}}{\alpha + \beta + v_{qd}}$$

Probability: working with an expected value

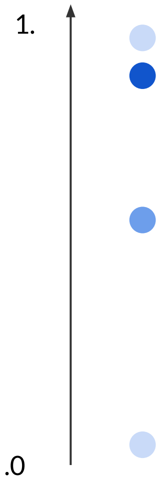
E[ctr]=0.1 | alpha=1, beta=9

clicks	views	ctr	y
20	100	0.20	0.19
0	100	0.00	0.01
0	1	0.00	0.09
1	1	1.00	0.18
99	100	0.99	0.91

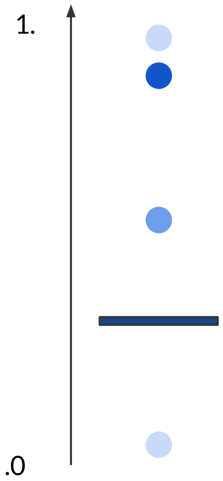
Shrinkage towards the expected value!

fewer views:
=> greater uncertainty
=> greater shrinkage
=> the more we rely on the expected value

Visual recap

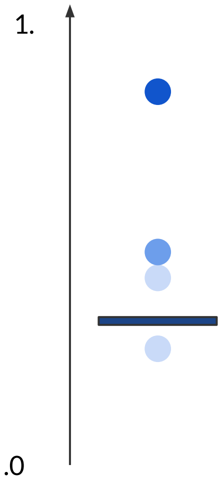


Observations (darker colours = more views)



$$E[ctr] = \frac{\alpha}{\alpha + \beta}$$

Expected value (Prior)



$$\frac{\alpha + c_{qd}}{\alpha + \beta + v_{qd}}$$

Expected value combined with observations (Posterior)

How can we find good values for alpha and beta?

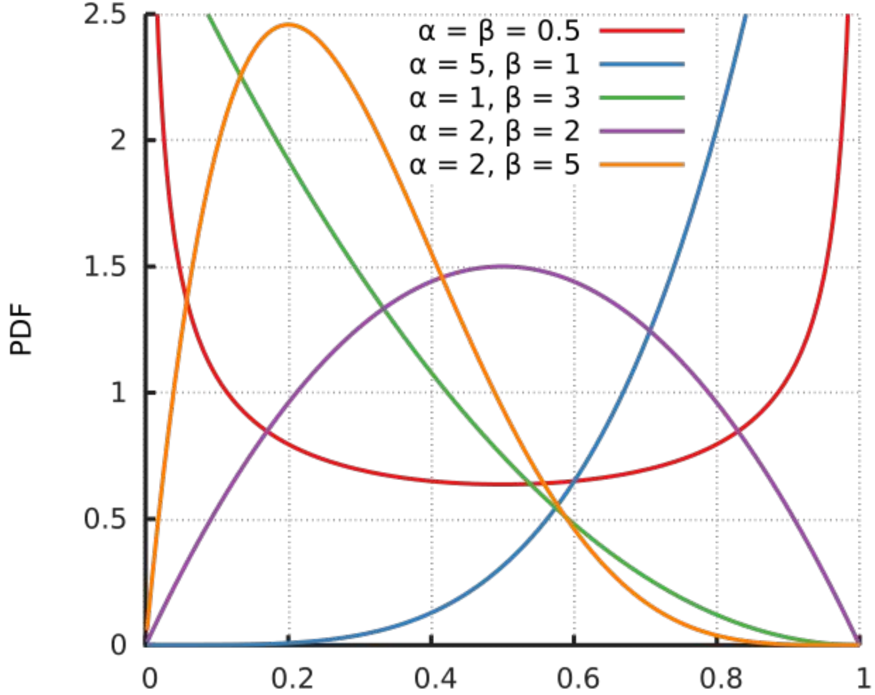
Shrinkage vs. choice of alpha and beta

$$y = \frac{\alpha + c_{qd}}{\alpha + \beta + v_{qd}}$$

The greater alpha and beta, the greater the shrinkage
=> we can model shrinkage together with E[ctr] by just using alpha and beta

clicks	views	ctr	α=0.01,β=0.09	α=1,β=9	α=100,β=900
20	100	0.20	0.20	0.19	0.11
0	100	0.00	0.00	0.01	0.09
0	1	0.00	0.01	0.09	0.10
1	1	1.00	0.92	0.18	0.10
99	100	0.99	0.99	0.91	0.18

Beta Distribution: parameters alpha and beta



Expected value of Beta(alpha,beta):

$$E[ctr] = \frac{\alpha}{\alpha + \beta}$$

https://en.wikipedia.org/wiki/Beta_distribution

Beta distribution: estimating alpha and beta

$$\alpha = \overline{ctr} \cdot (\alpha + \beta)$$

$$\beta = (1 - \overline{ctr}) \cdot (\alpha + \beta)$$

$$\alpha + \beta = \frac{\overline{ctr} \cdot (1 - \overline{ctr})}{s^2} - 1$$

Estimation based on sample mean (CTR) and varian
s^2

$$\text{if } \overline{ctr} \cdot (1 - \overline{ctr}) > s^2$$

Great intuition: (alpha + beta) ~ $1/s^2$

lower variance ~ greater (alpha+beta)

Intuition: the greater certainty (via variance), the greater the shrinkage, the more we trust our expected ctr value, the more views/observations we need to diverge from it

alpha ~ probability of success, beta ~ probability of failure

We can write Beta(mean,variance) instead of Beta(alpha,beta)

Welcome to Bayesian Modelling

Bayes' Theorem with a 'Beta prior'

$$\frac{\alpha}{\alpha+\beta}$$

Prior

Likelihood function

$$P(C=1|D) = \frac{P(C=1)\cdot P(D|C=1)}{P(D)}$$

Posterior

$$\frac{\alpha + c_{qd}}{\alpha+\beta+v_{qd}}$$

We can solve click probability and certainty, but…

How can we eliminate position bias and other contexts from our model?

Context: dealing with position bias

COEC - Clicks over expected clicks

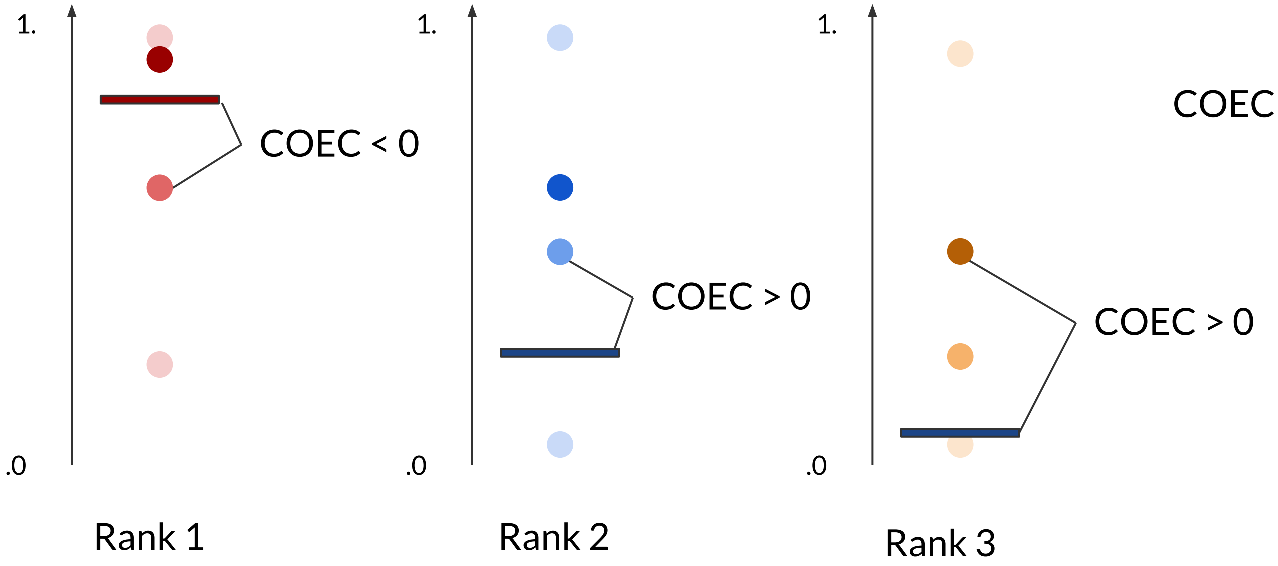
$$\frac{\sum_{r=1}^{N} c_{qd_r}}{\sum_{r=1}^{N} \overline{ctr}_r \cdot v_{qd_r}}$$

r: rank at which clicks and views were observed

$$\frac{ctr_{qd_r}}{\overline{ctr}_r}$$

Same as above if query-doc pair was only observed at a single rank

Position bias: COEC



$$\text{COEC} = \frac{ctr_{qd_r}}{ctr_r}$$

Context: dealing with position bias

COEC - Clicks over expected clicks

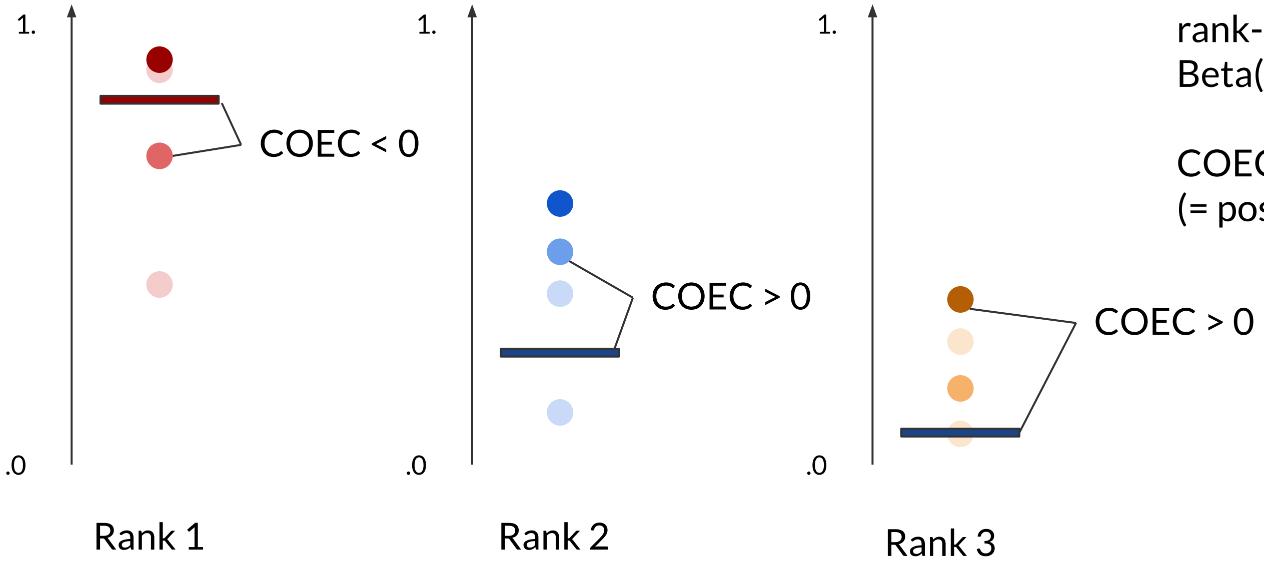
$$\frac{\alpha + c_{qd}}{\alpha + \beta + v_{qd}}$$

$$\frac{ctr_{qd_r}}{\overline{ctr}_r}$$

$$E[ctr] = \frac{\alpha}{\alpha + \beta}$$

Same as above if query-doc pair was only observed at a single rank

Position bias: COEC after shrinkage



rank-specific priors
Beta(alpha$_r$,beta$_r$)

COEC still works
(= posterior/prior)

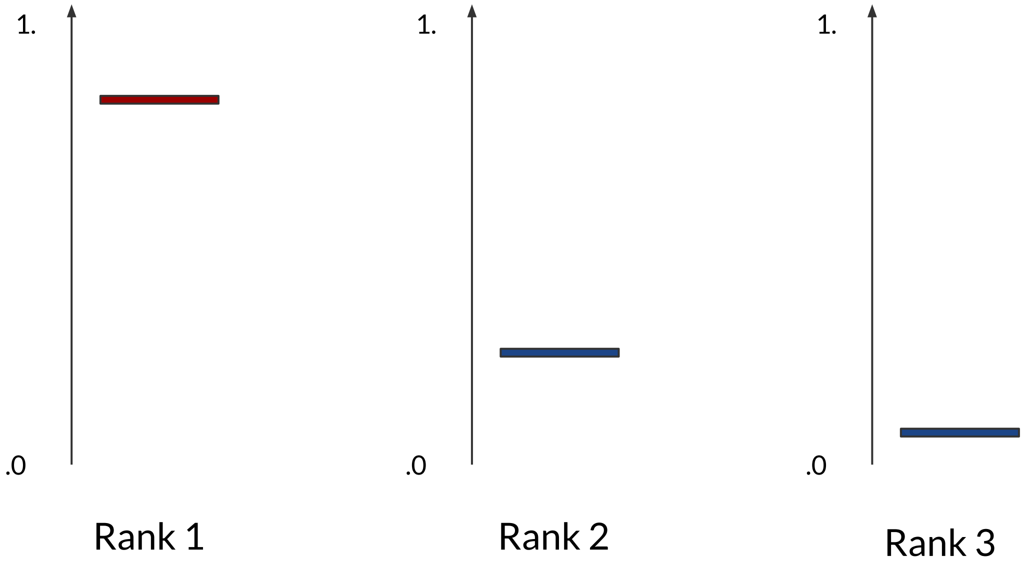
Position bias: COEC after shrinkage

Normalise by expected ctr for a given rank => great intuition!

We can combine COEC with a beta prior (shrinkage) to deal with ‘certainty’. => We still need to see how we can deal with documents that occur at more than one position

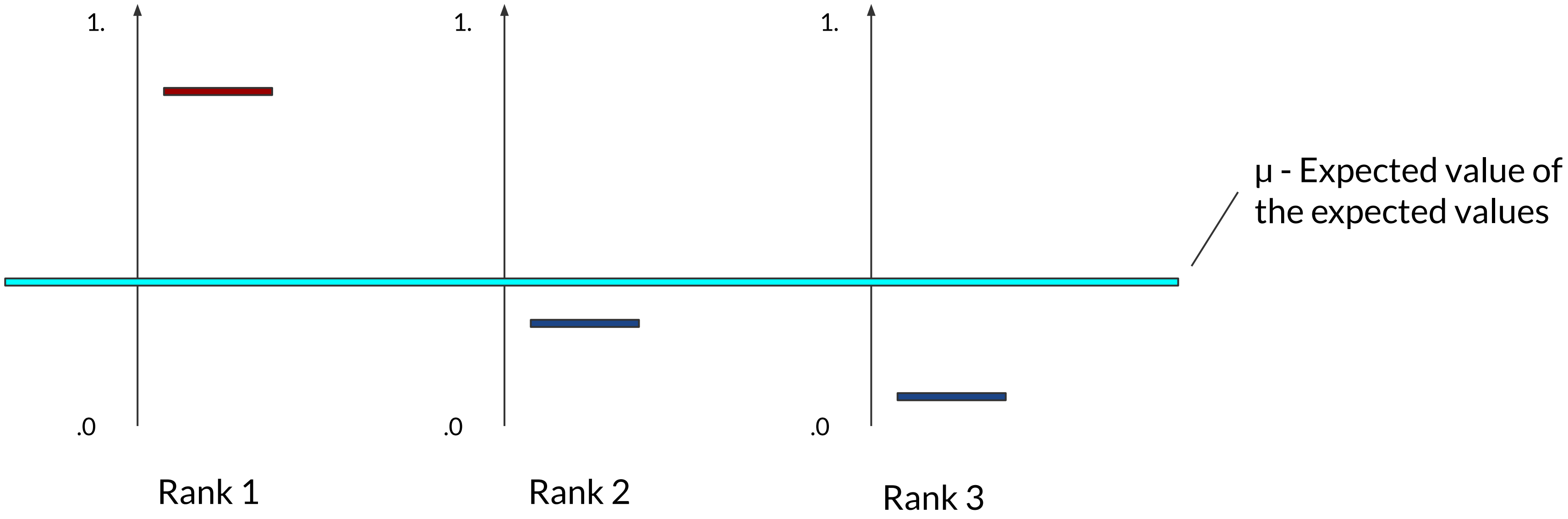
BUT: COEC not only cancels out position bias but als the impact of the ranking that was applied the search engine when we collected our observations => we’ll deal with this!

Position bias and impact of ranking

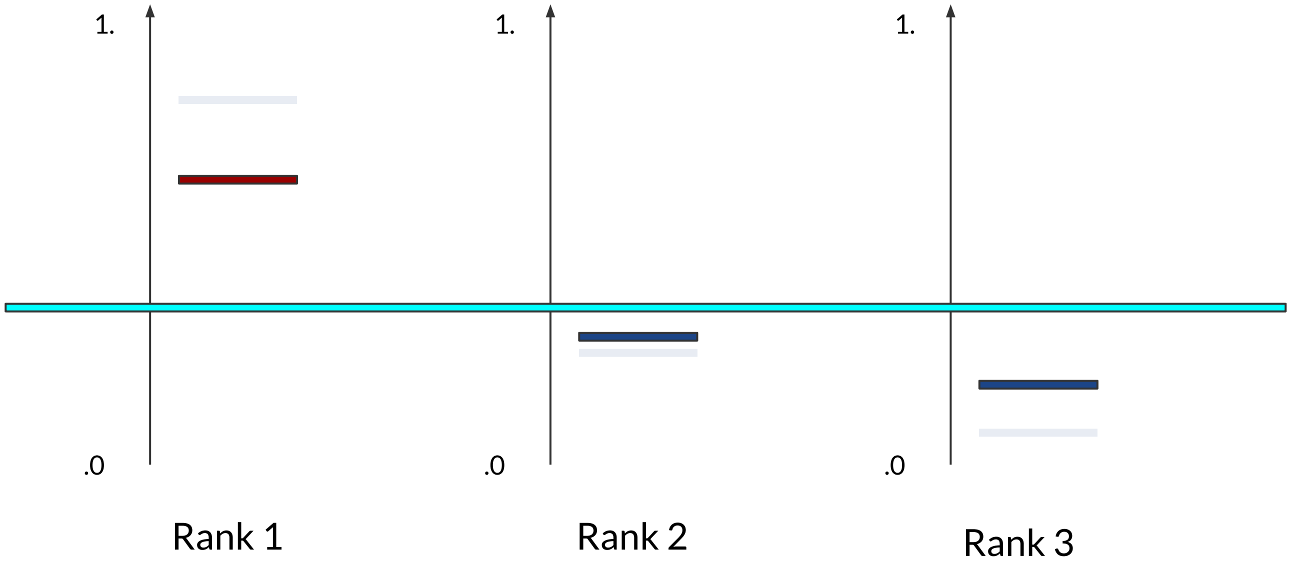


Expected value per rank

Expected value of the per-rank means



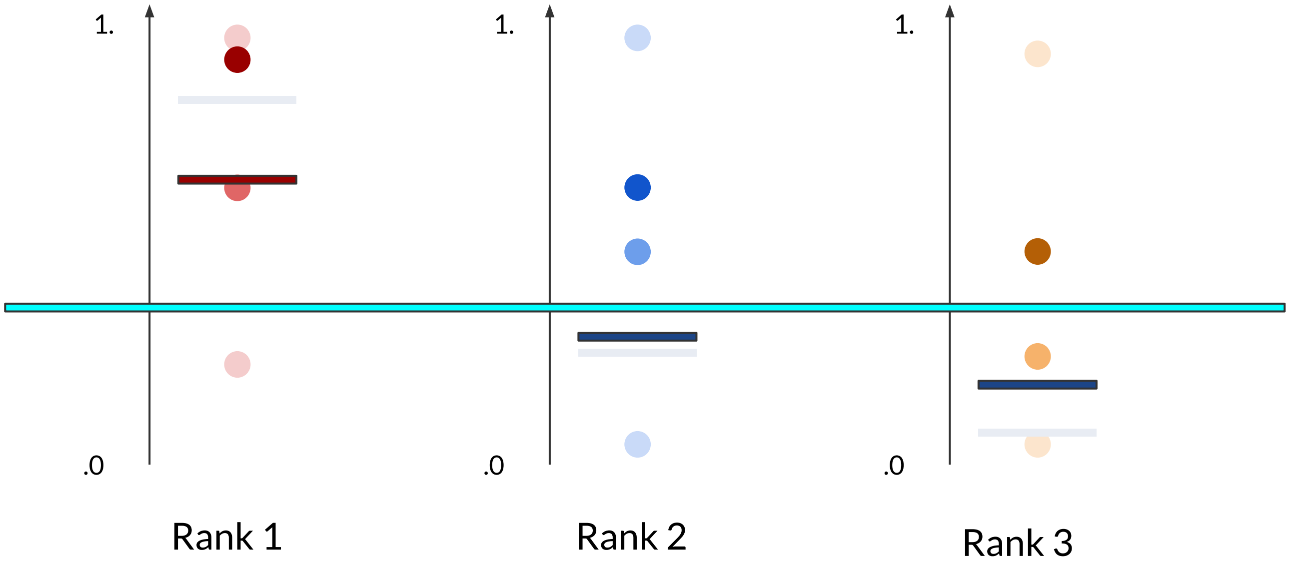
Beta distribution of expected mean CTRs



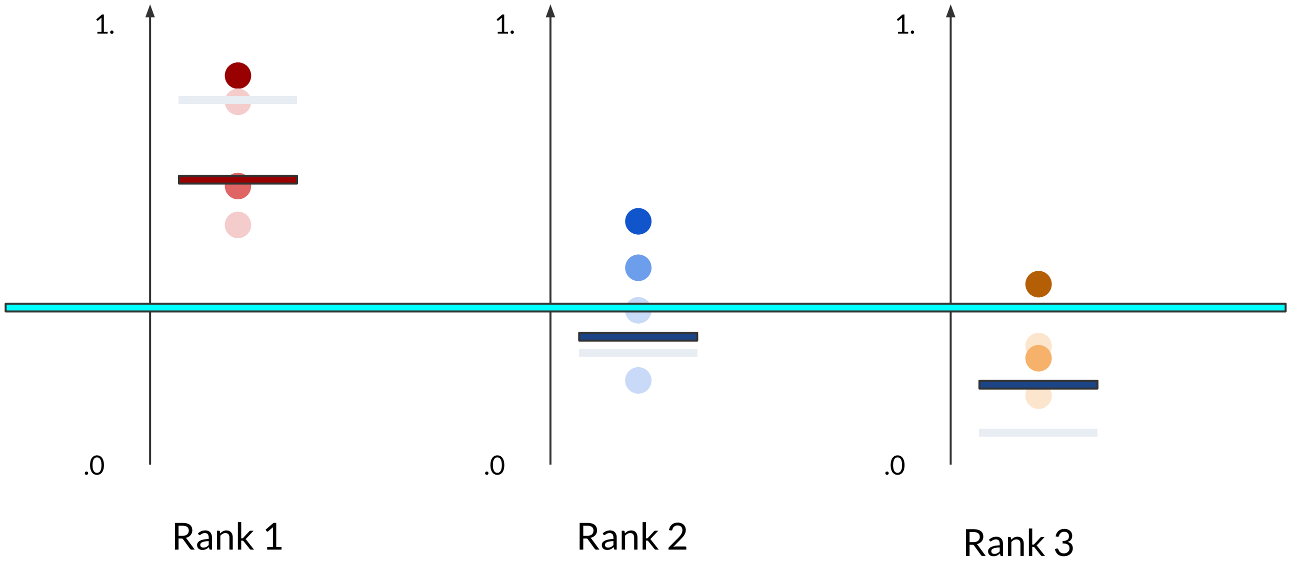
μ,σ^2 - Expected value and variance of the expected values

Posterior distribution using Beta(μ,σ^2)

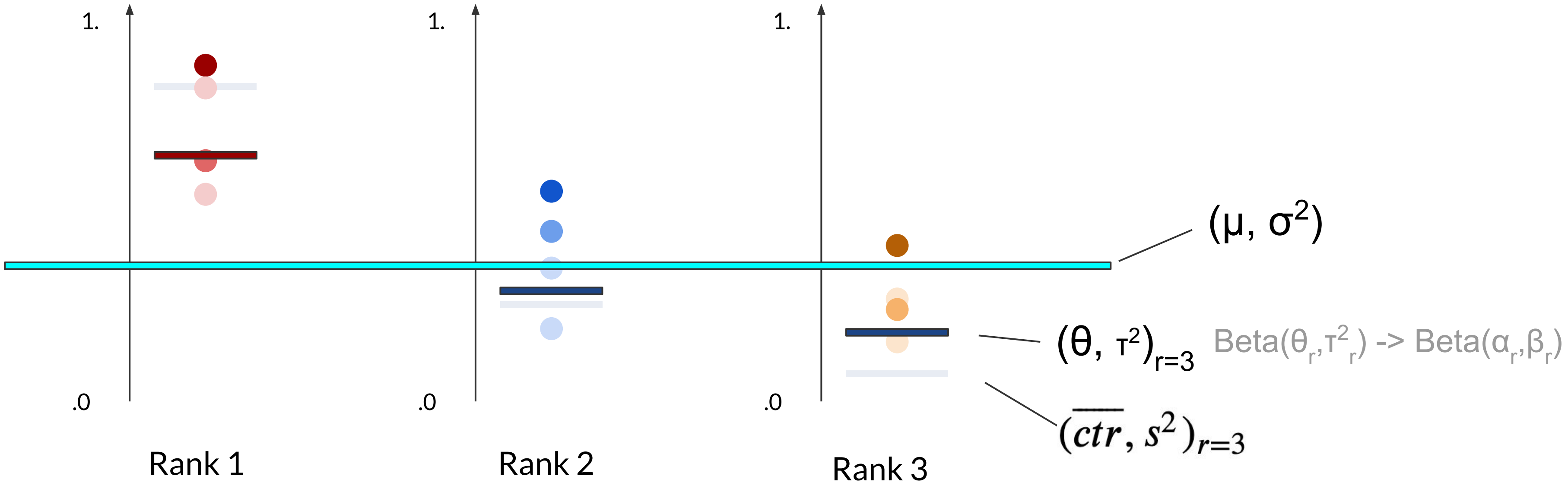
Multi-level modelling: 1) expected mean CTR



Multi-level modelling: 2) posterior CTR

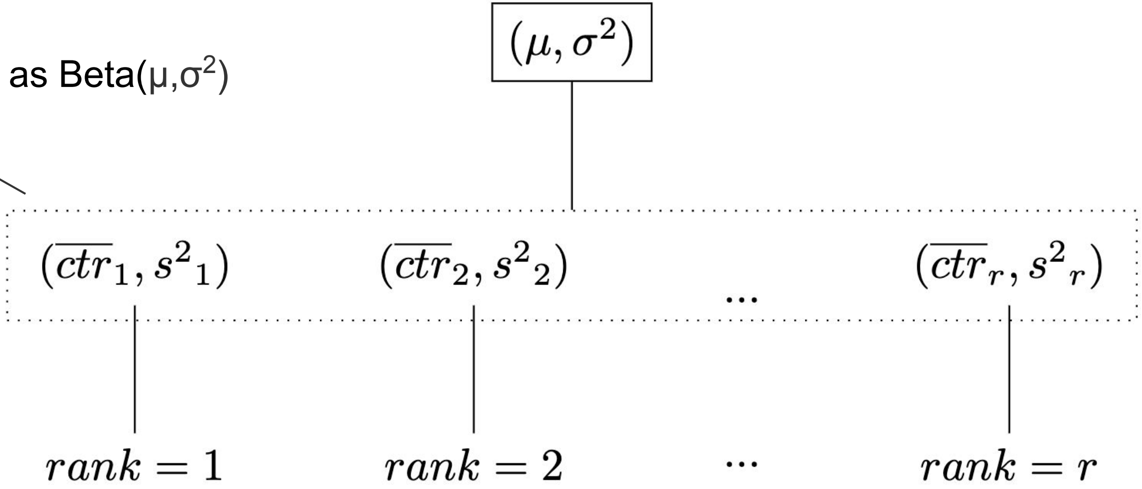


Multi-level modelling



Partial pooling / hierarchical model

Rank-specific priors modelled as Beta(μ,σ^2)



Partial pooling: each rank has its prior parameters $(\overline{ctr}_r, s^2_r)$ - like in the unpooled approach. Now we assume a beta distribution of these priors with parameters μ (mean) and σ^2 (variance)

Partial pooling / hierarchical Bayesian model

‘Prior of priors’ Per rank (unpooled)

$$\theta_r = \frac{\frac{1}{\sigma^2} \cdot \mu + \frac{1}{{s^2}_r} \cdot \overline{ctr}_r}{\frac{1}{\sigma^2} + \frac{1}{{s^2}_r}}$$

$$\tau^2{}_r = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{{s^2}_r}}$$

Prior at rank r: Beta(θ_r,τ^2_r)

remember that we can calculate (θ_r,τ^2_r) -> (α_r,β_r)

Weighted average of the pooled mean μ and the rank-specific mean ctr

Using variance as weight! - If the ctr varies a lot at rank r - the pooled mean becomes more important

Keeps global relevance information

Models position bias

Estimation of μ and σ^2:
- bootstrapping
- samples weighted by views at r
- Gelman et al.: use 99th percentile variance

Hierarchical Bayesian model

'Prior of priors'   Per rank (unpooled)

$$\theta_r = \frac{\frac{1}{\sigma^2} \cdot \mu + \frac{1}{s^2{}_r} \cdot \overline{ctr}_r}{\frac{1}{\sigma^2} + \frac{1}{s^2{}_r}}$$

$$\tau^2{}_r = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{s^2{}_r}}$$

Prior at rank r: Beta(θ_r,$\tau^2{}_r$)

remember that we can calculate (θ_r,$\tau^2{}_r$) -> (α_r,β_r)

Posterior

$$P(C_{qd_r} = 1|D) = \frac{\alpha_r + c_{qdr}}{\alpha_r + \beta_r + v_{qdr}}$$

$$Judgment_{qd} = \frac{P(C_{qd_r} = 1|D)}{\theta_r}$$

Aggregating across ranks

$$w_{qd_r} = \frac{v_{qd_r}}{v_{qd}} \cdot \frac{1}{{s^2}_r}$$

$$\theta_{qd} = \frac{\frac{1}{\sigma^2} \cdot \mu + \sum_{r=1}^{N} w_{qd_r} \overline{ctr}_r}{\frac{1}{\sigma^2} + \sum_{r=1}^{N} w_{qd_r}}$$

$${\tau^2}_{qd} = \frac{1}{\frac{1}{\sigma^2} + \sum_{r=1}^{N} w_{qd_r}}$$

If a query-doc pair was observed at multiple ranks, use the views-weighted rank-specific parameters

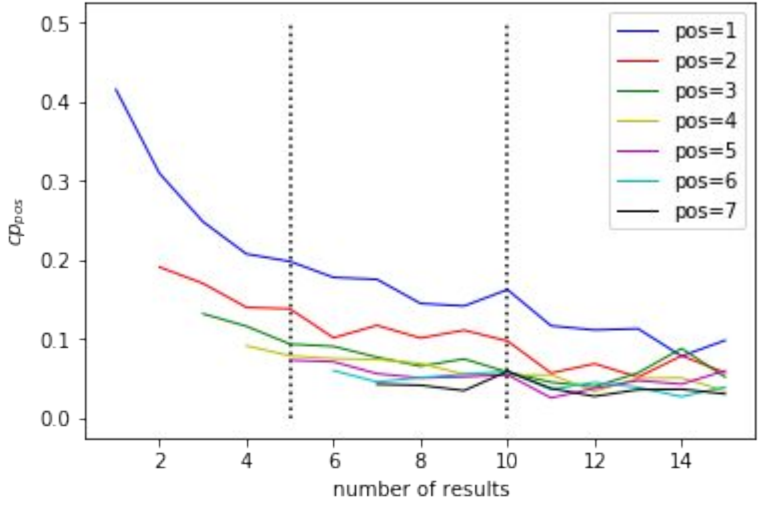
Do not attempt to aggregate the posteriors instead of the priors (it took me 1.5 years to figure this out ;-)) -> think: shrinkage per rank

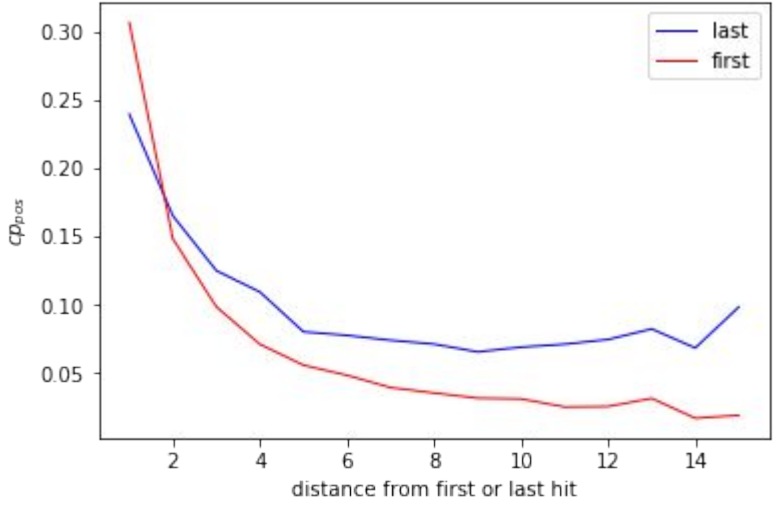
Posterior:

$$P(C_{qd} = 1|D) = \frac{\alpha_{qd} + c_{qd}}{\alpha_{qd} + \beta_{qd} + v_{qd}}$$

From ranks to contexts

User do not (always) scan results sequentially or just stop at a good result

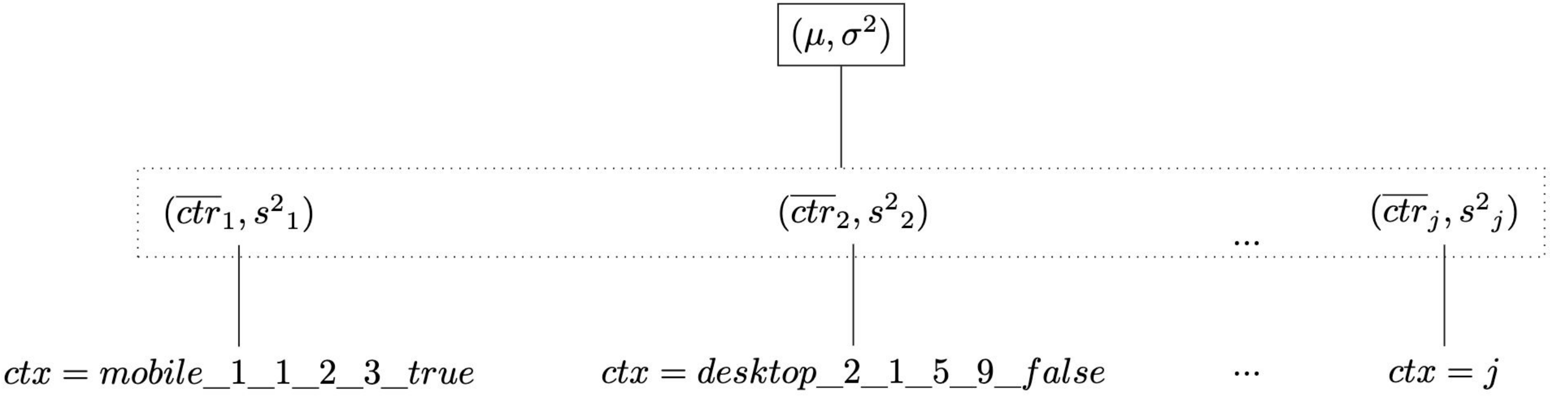




cp=ctr, data taken from 5-per-row grid layout site

From ranks to contexts

- Do not assume any relationship between ranks - results are not processed in a known sequence
- Rank r becomes Context j



desktop_2_1_5_9_false: device=desktop, page=2, results-per-row=5, position=5, results-shown-on-page=9, has-marketing-label=false

Using contexts

$$w_{qd_j} = \frac{v_{qd_j}}{v_{qd}} \cdot \frac{1}{{s^2}_j}$$

$$\theta_{qd} = \frac{\frac{1}{\sigma^2} \cdot \mu + \sum_{j=1}^{N} w_{qd_j} \overline{ctr_j}}{\frac{1}{\sigma^2} + \sum_{j=1}^{N} w_{qd_j}}$$

$${\tau^2}_{qd} = \frac{1}{\frac{1}{\sigma^2} + \sum_{j=1}^{N} w_{qd_j}}$$

Simply replace rank r with context j

Note that we don't assume an order in which the user would scan the results! (Think of grid layout in e-commerce)

If you don't have samples to calculate a prior for a given context, you can fall back to using just (μ, σ^2)

Posterior

$$P(C_{qd} = 1|D) = \frac{\alpha_{qd} + c_{qd}}{\alpha_{qd} + \beta_{qd} + v_{qd}}$$

$$Judgment_{qd} = \frac{P(C_{qd}{=}1|D)}{\theta_{qd}}$$

Combining events

Calculate judgments using separate hierarchical Bayesian Models, multiply judgments, you may want to weight event types

$$Judgment_{qd} = \frac{P(C_{qd}=1|D)}{\theta_{clicks_{qd}}}^{w_1} \frac{P(O_{qd}=1|D)}{\theta_{orders_{qd}}}^{w_2}$$

Thank you!

René Kriegler
rkriegler@o19s.com

OpenSource Connections
https://o19s.com

For more e-commerce search fun:

Join us! - https://o19s.com/about-us/careers/

A special thanks to Maximilian Kusterer who introduced me to the idea of using a beta prior.