Sabre

Product Recommendations with Multi-Armed Bandit

Hunkar Toyoglu
Richard Ratliff

AGIFORS RM Study Group
May 08, 2019
Panama City, Panama

AGIFORS
Airline Group of the International Federation of Operational Research Societies

Agenda

- Intelligent retailing
- Multi-Armed Bandit with Thompson Sampling
- Retail Product Recommendations
- Next Steps
- Summary
- Q&A

INTELLIGENT RETAILING

Achieve revenue maximization and provide differentiated experiences

Traveler expectations are changing

66%
of mobile travelers are comfortable planning and booking a vacation only with a smartphone

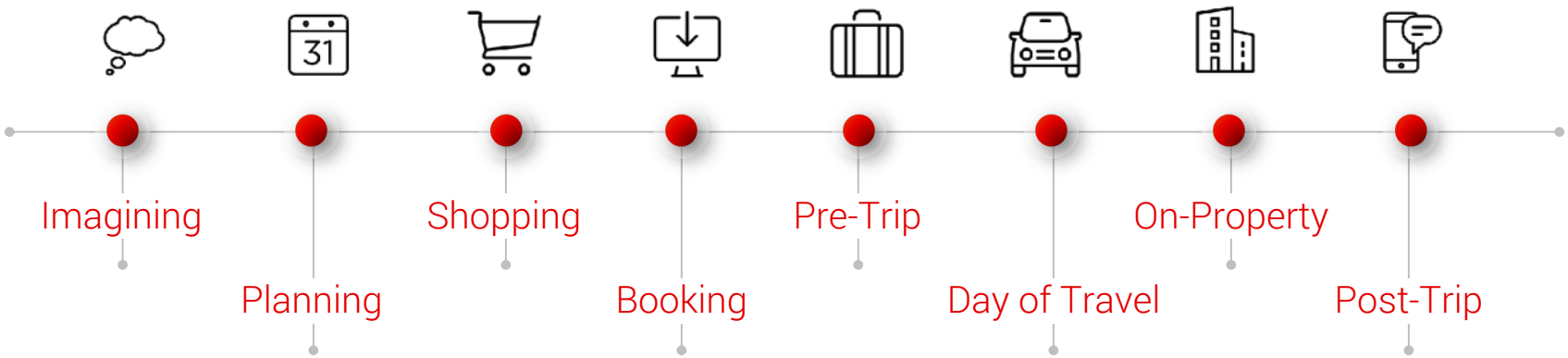
60%
of U.S. travelers would consider an impulse trip based on a good hotel or flight deal

57%
of U.S. travelers feel that brands should tailor their information based on personal preferences or past behaviors

94%
of global travelers want to pick and choose their own travel add-ons to create a personalized experience

Sources:
Phocuswright US Mobile Traveler, 2017
Gartner, Edison Research, 2017
The Global Traveler Research Report, Sabre, 2017
Google/Phocuswright Travel Study, 2017

Increase revenue while enhancing guest experience with solutions that span full journey



Digital >> Data

+

Mobile >> Touchpoints

[anytime, anywhere]

+

AI >> Automation

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Travel industry is laser-focused on becoming better retailers

Changes in the market
Evolving customer behaviors
Emerging technologies

Relevant, personalized offers

Ready to intelligently retail?

What are the needs of the customer?

What intelligent infrastructure is needed to meet the company needs?

How to provide the best retailing options and win against competition?

MULTI-ARMED BANDIT WITH THOMPSON SAMPLING

Exploration vs Exploitation

An intelligent platform for retailing

- Select an ordered list of products to display (MAB-TS)
- Infinite SKUs - anything store
  - things (robe, blanket, ...)
  - policies (early check in, late check out, ...)
  - services (airport transport, flower delivery, ...)
  - experiences (massage, bike rental, ...)
- SKUs are not attached to the room

Living Spaces

Work Spaces

Experiences

Merchandise

Services

Amenities

Policies

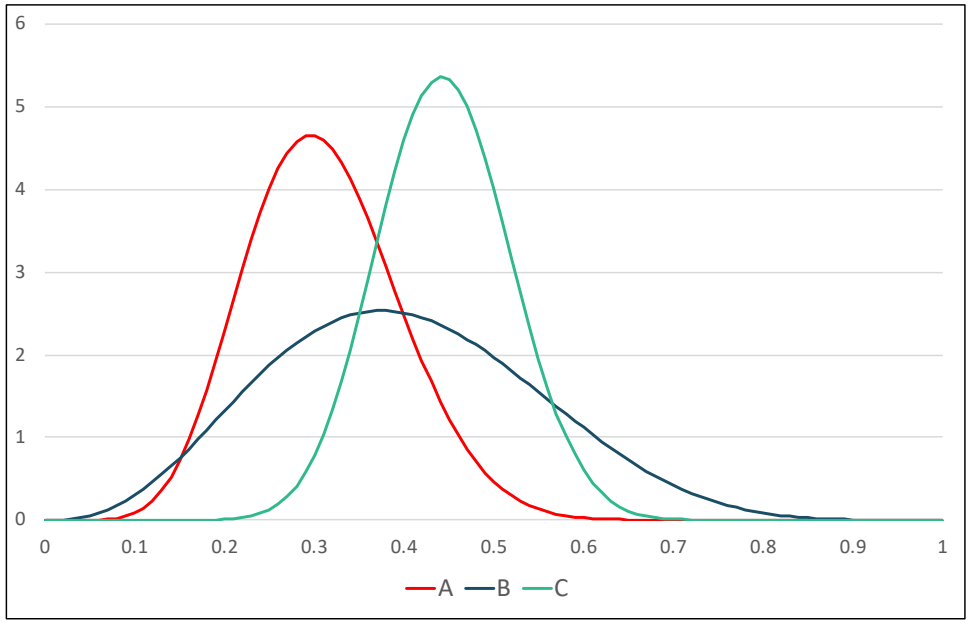
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Stochastic Multi Armed Bandit Problem

- A slot machine with multiple arms
- Each arm produces an i.i.d. random reward independent of the past and other arms
- Parameters of the reward distribution are fixed but unknown
- A reward can be observed only after playing an arm

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Trade-off: Best reward or more information



Posterior distributions of mean rewards observed in previous time periods

- **Exploiting** arms that yielded high rewards in the past for immediate rewards
- **Exploring** other arms to gather more information for long term rewards
- Develop a sequential strategy to balance this tradeoff and max the total expected reward or min total expected regret

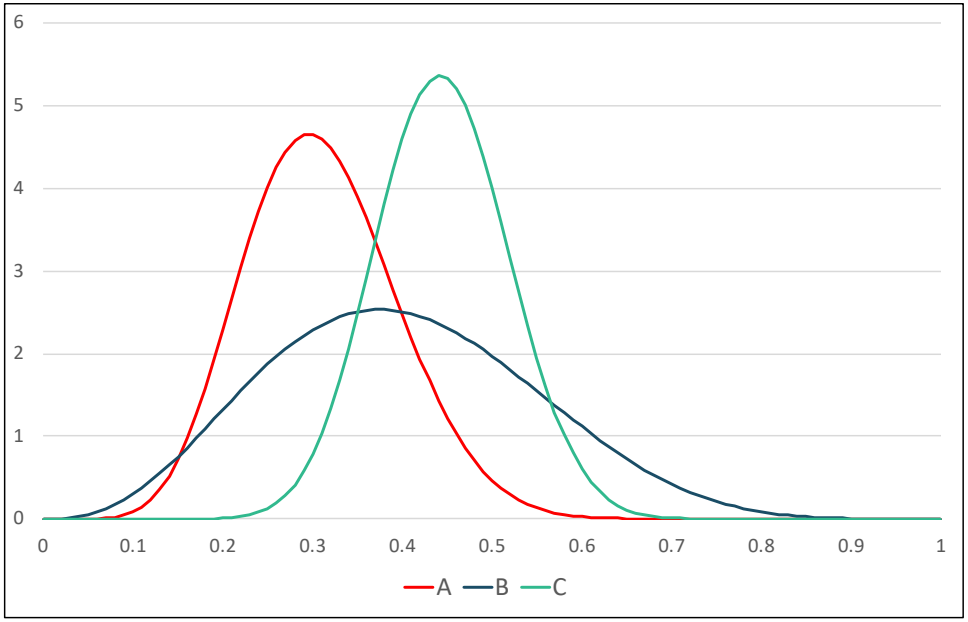
➢ Online learning

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Bernoulli Bandit

- K arms/actions
- Each action produces either:
  - a success (1) with Q
  - a failure (0) with 1 - Q
- max total # of successes
- Q is unknown but fixed/constant
- Q can be learned only by experimenting

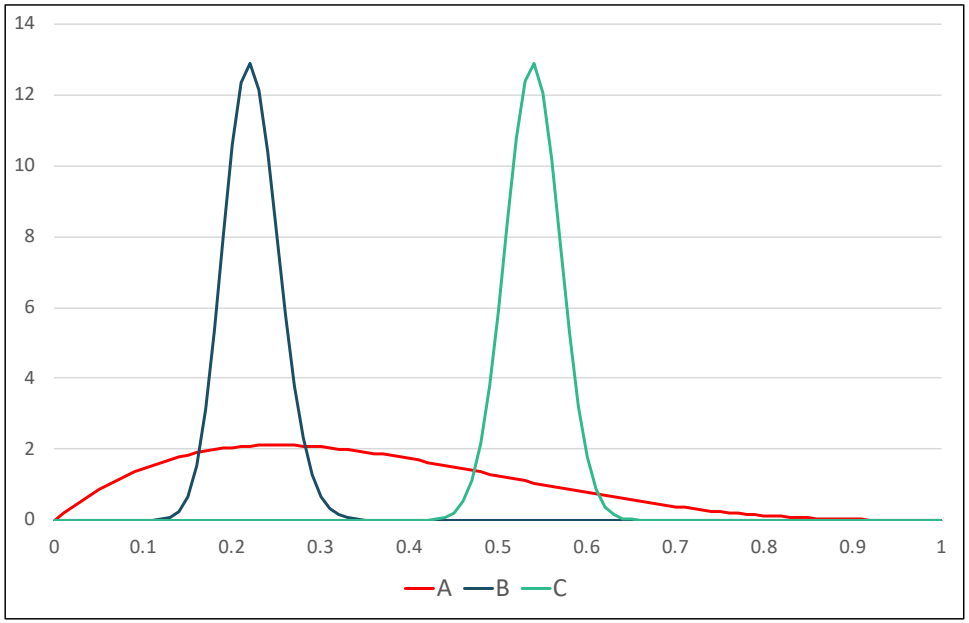
- Naive approach:
  - Allocate a fixed time period to exploration
  - Exploit in other time periods



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Greedy algorithm

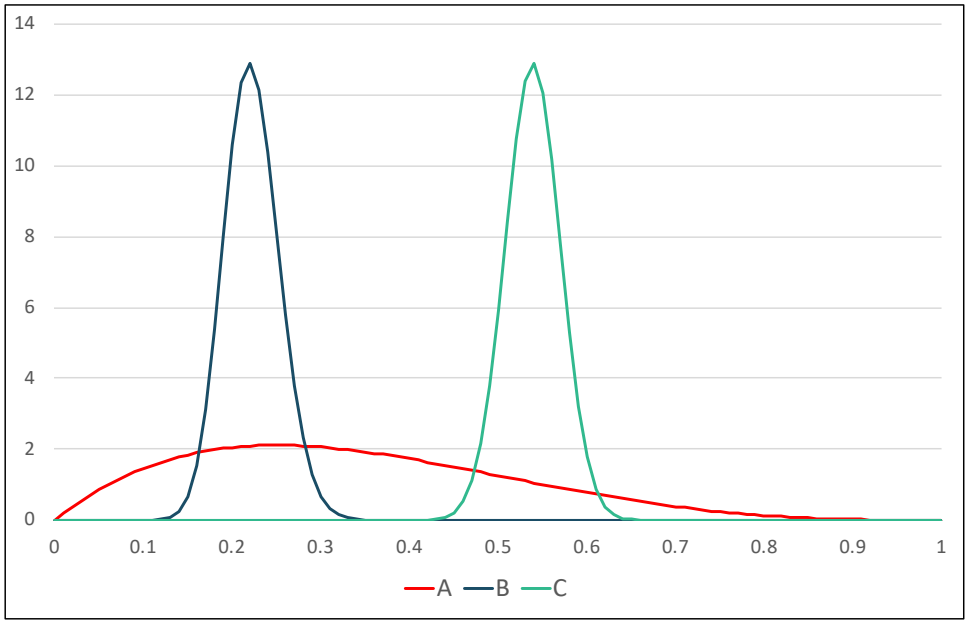
- Select an action to maximize immediate reward (100% exploit)
- Select C due to max E[Q]
- Avoid inferior action B
- As uncertainty of C gets smaller, it will continue to be selected
- Fails to recognize the uncertainty of A and does not actively explore it



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

ε-greedy exploration

- Select;
  - the greedy action with probability $1 - \varepsilon$
  - an action uniformly at random otherwise
- Fails to eliminate unpromising actions, hence does not explore promising actions more
- Should eliminate B due to no chance of being optimal

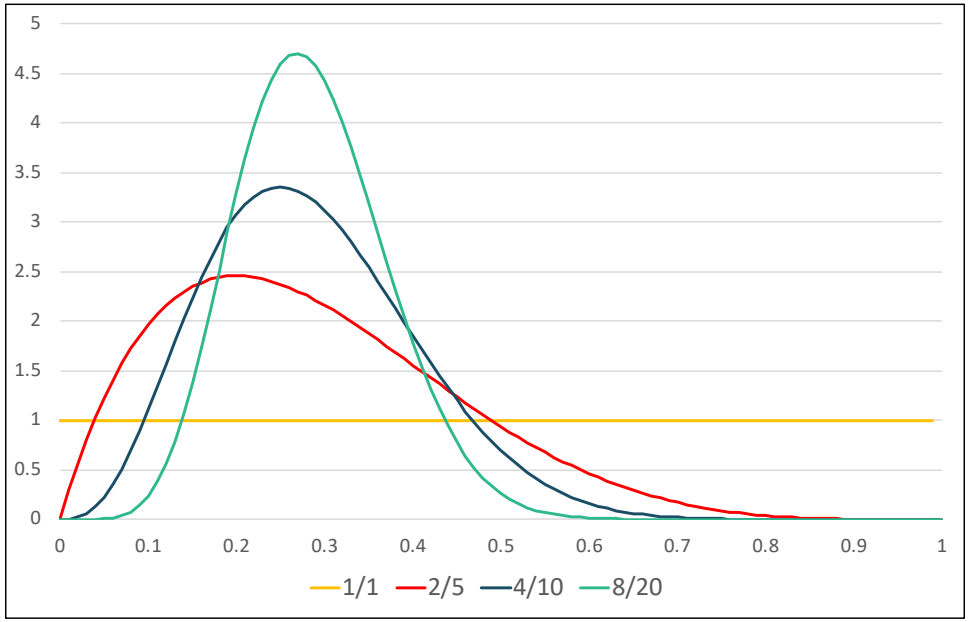


confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Thompson sampling
(posterior sampling / probability matching)

- TS is a Bayesian approach
- Assume a prior distribution over Q of each action
- Let priors of Q are beta(α , $\mathcal{B}$) where
  - α = # of successes
  - $\mathcal{B}$ = # of failures
- $Q = \alpha / (\alpha + \mathcal{B})$
- Posteriors of Q are also beta(α+1 , $\mathcal{B}$) or beta(α , $\mathcal{B}$+1)

Pick an action according to its posterior probability of being the best action



Distribution gets more concentrated as observations grow

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Thompson Sampling vs Greedy

Thompson Sampling

- Select an action according to sampling of Q from its posterior distribution
- for k in K
  - Sample $Q_k \sim \text{beta}(\alpha_k, \mathcal{B}_k)$
  - Action $\text{argmax}_k\, Q_k$
  - Observe reward
  - Update posterior: $(\alpha_k, \mathcal{B}_k) \leftarrow (\alpha_k+1, \mathcal{B}_k+1)$
- end for

Greedy

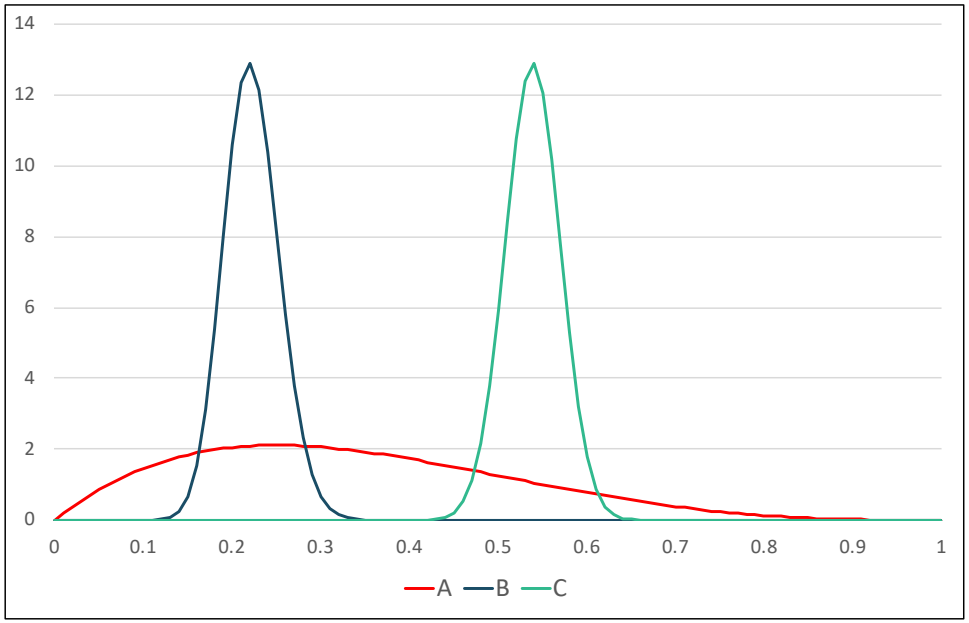
- Select an action according to estimation of Q from its posterior distribution
- for k in K
  - Estimate $Q_k = \alpha_k / (\alpha_k + \mathcal{B}_k)$
  - Action $\text{argmax}_k\, Q_k$
  - Observe reward
  - Update posterior: $(\alpha_k, \mathcal{B}_k) \leftarrow (\alpha_k+1, \mathcal{B}_k+1)$
- end for

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Thompson sampling

- Greedy only exploits C but does not explore potential value of A
- ε-greedy explores both A and B equally although exploring B is futile
- TS samples actions based on their probability of being best
- Due to randomization TS is unlikely to get stuck with an inferior action

p(C is best action) >
p(A is best action) >
p(B is best action) = 0



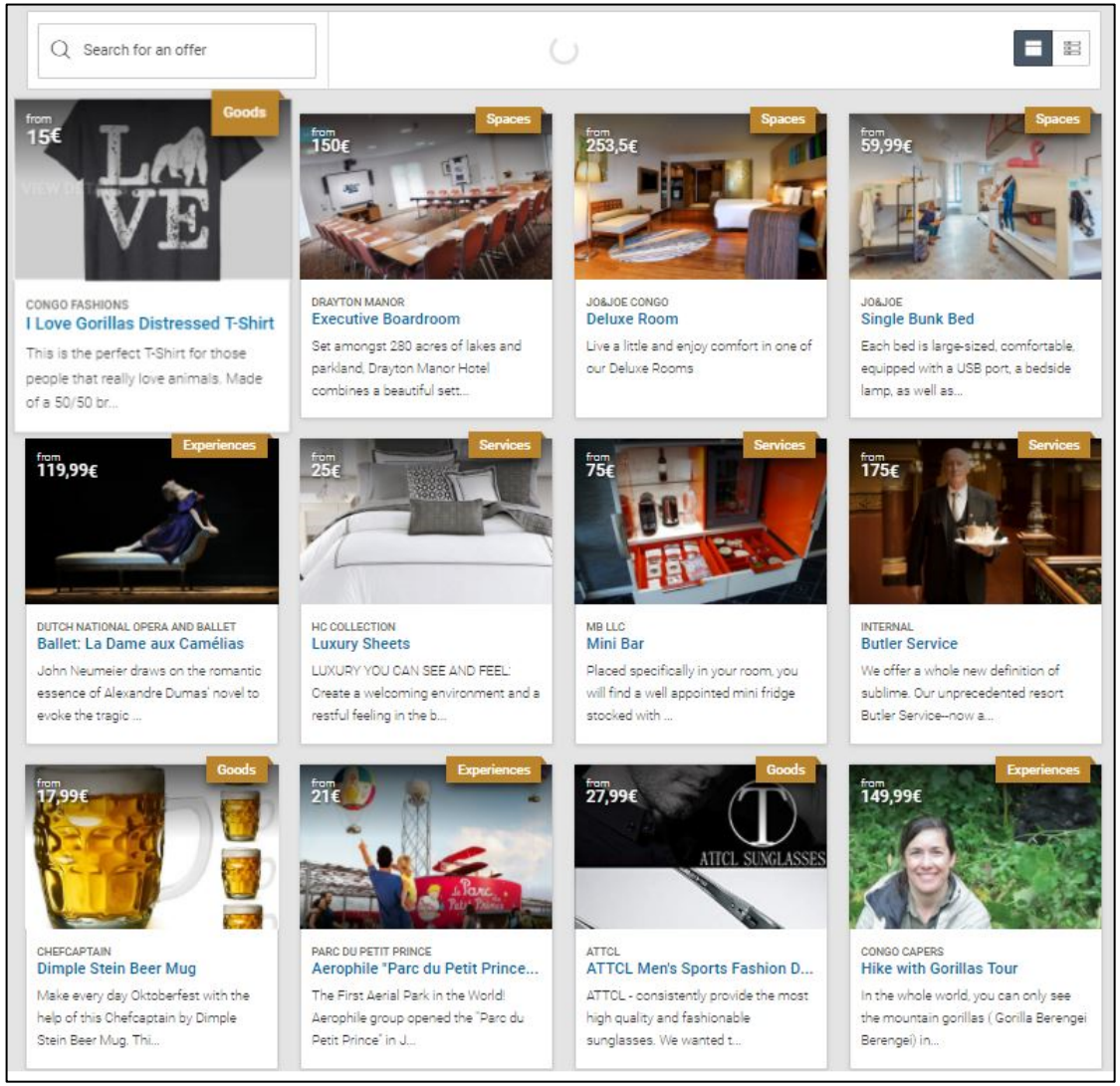
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

RETAIL PRODUCT RECOMMEN-DATIONS

Optimize which products to show visitors

Retail product recommendations

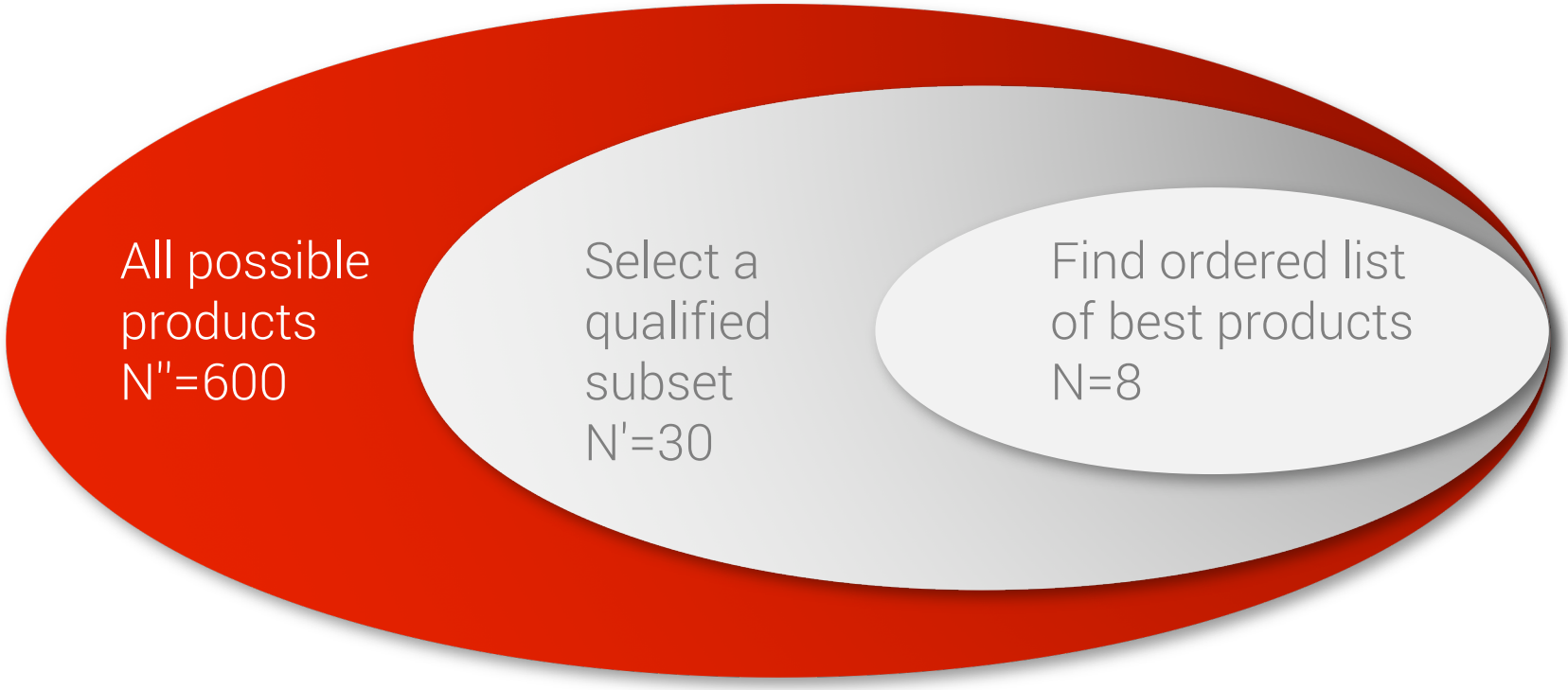
- **Select the best content to display on website to the next visitor**
  - Which specific products are best displayed for this customer segment?
  - What is the best display order of those products?
- **Reward is gained if the user click on a content**



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

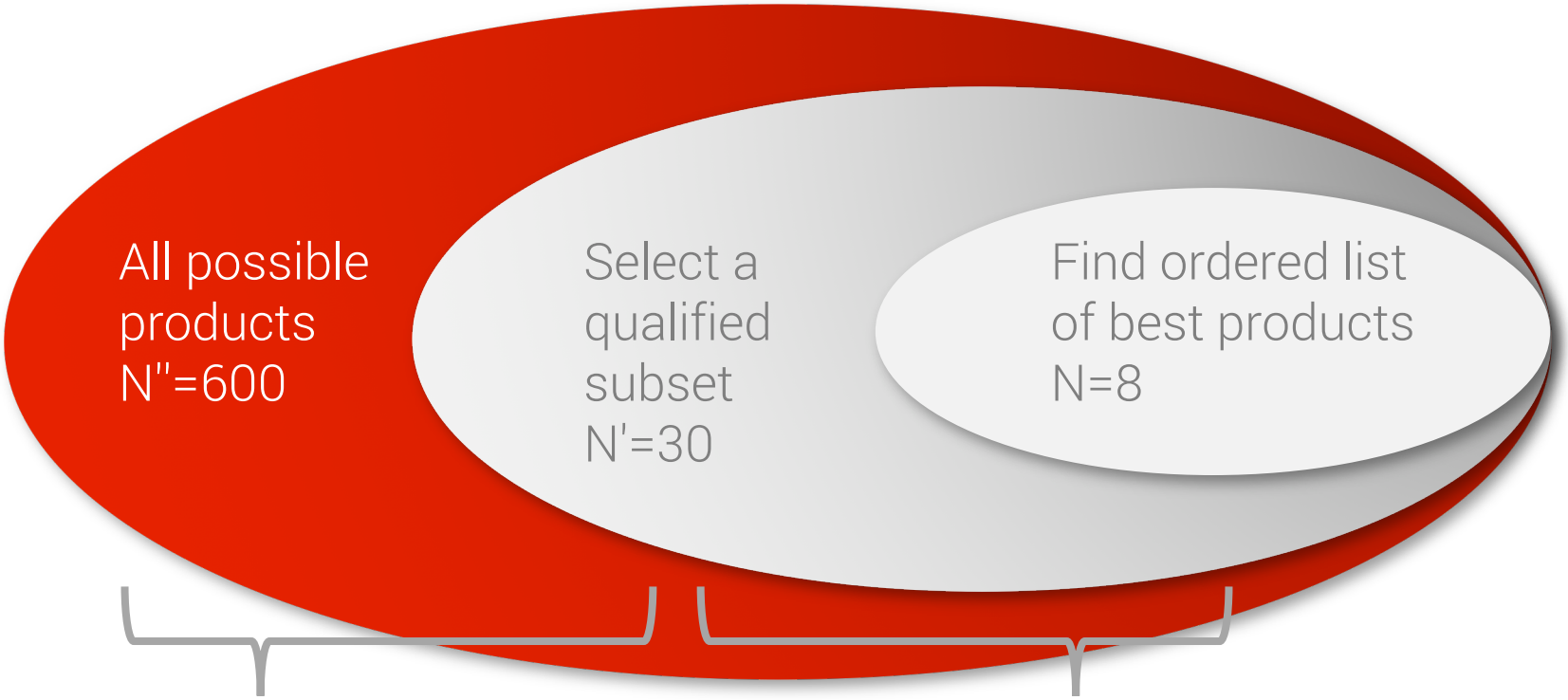
Recommender processing - 1

Objective is to display the best 8 items in the right order from an original set of 600 items given a context



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Recommender processing - 2



Use pre-processing business logic or sales correlation to choose logical candidates

Display [3, 2, 5, 1, 18, 6, 4, 27]
Don't display the remaining candidates

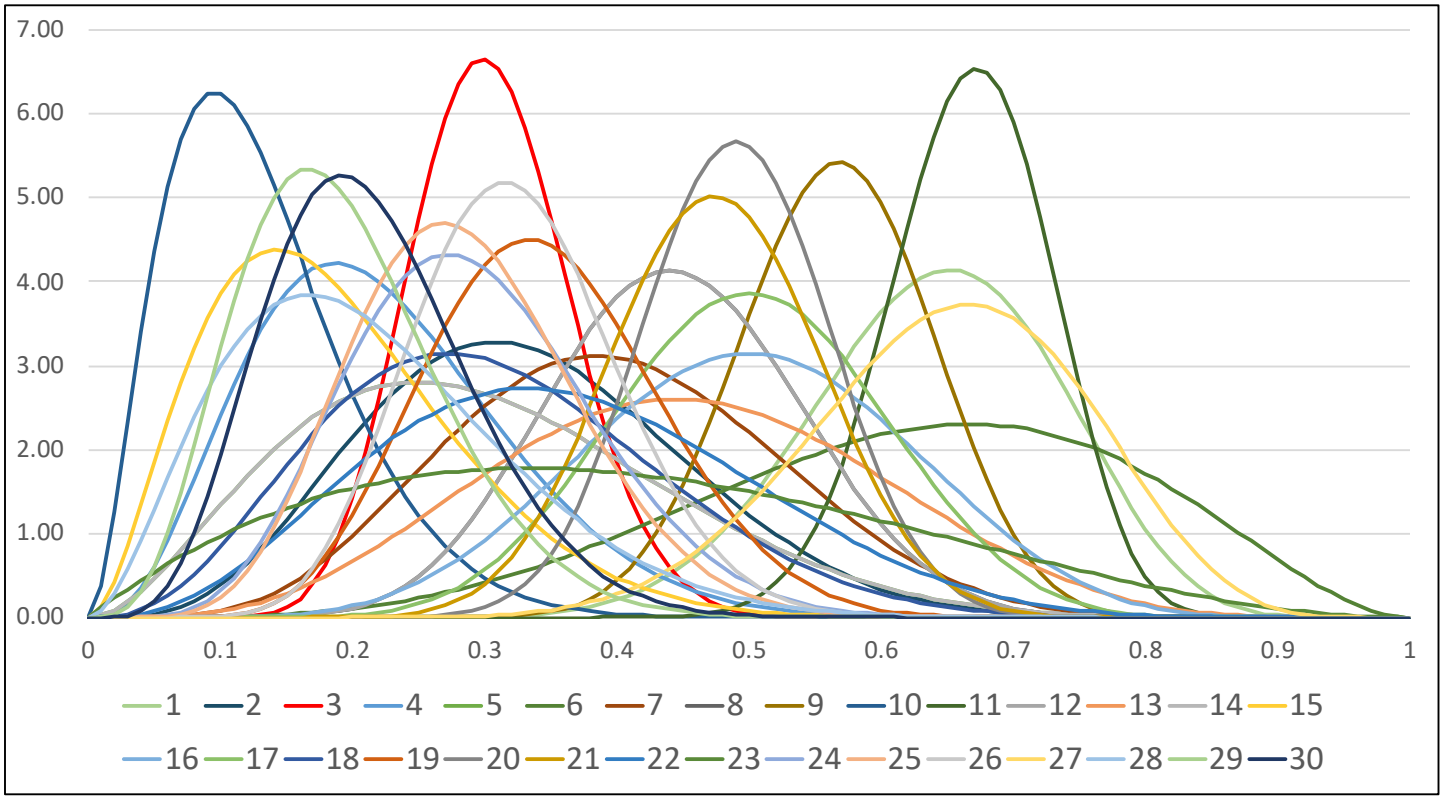
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Posterior distributions with beta(α_k , $\mathcal{B}_k$)

N'=(1,2,3,...,30)

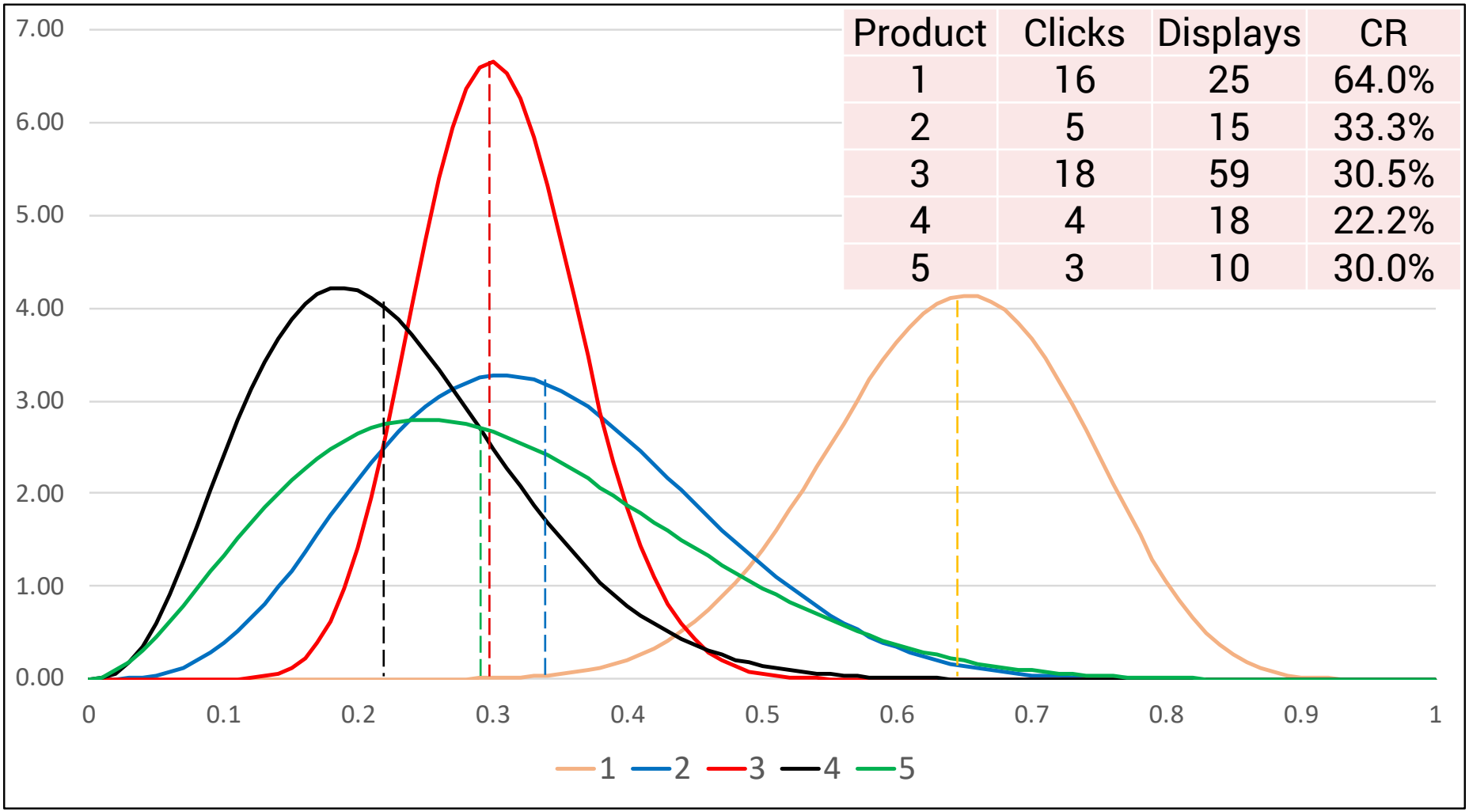
α = # sales
$\mathcal{B}$ = # !sales

N=()



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

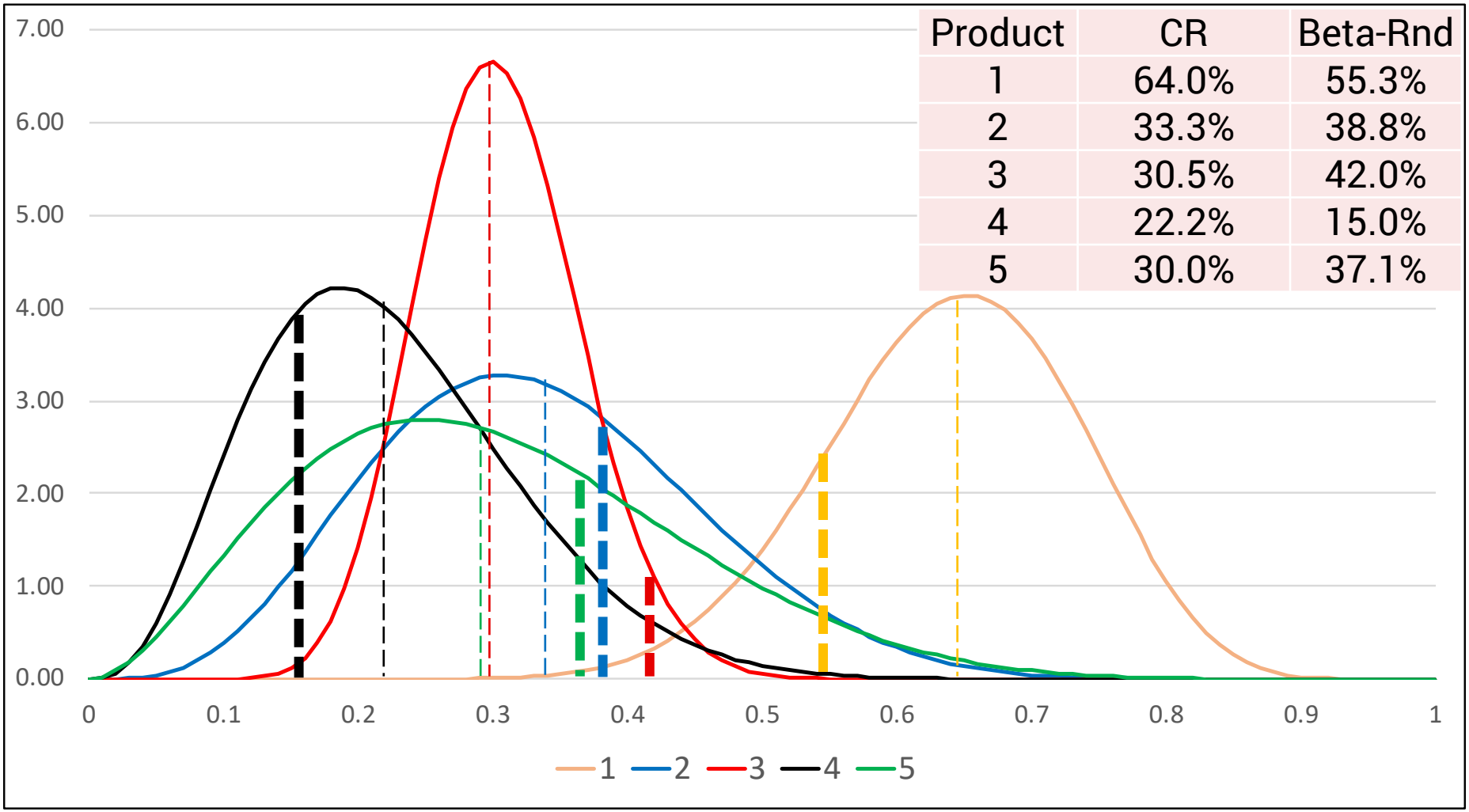
Display selection – Conversion Rate



Product	Clicks	Displays	CR
1	16	25	64.0%
2	5	15	33.3%
3	18	59	30.5%
4	4	18	22.2%
5	3	10	30.0%

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

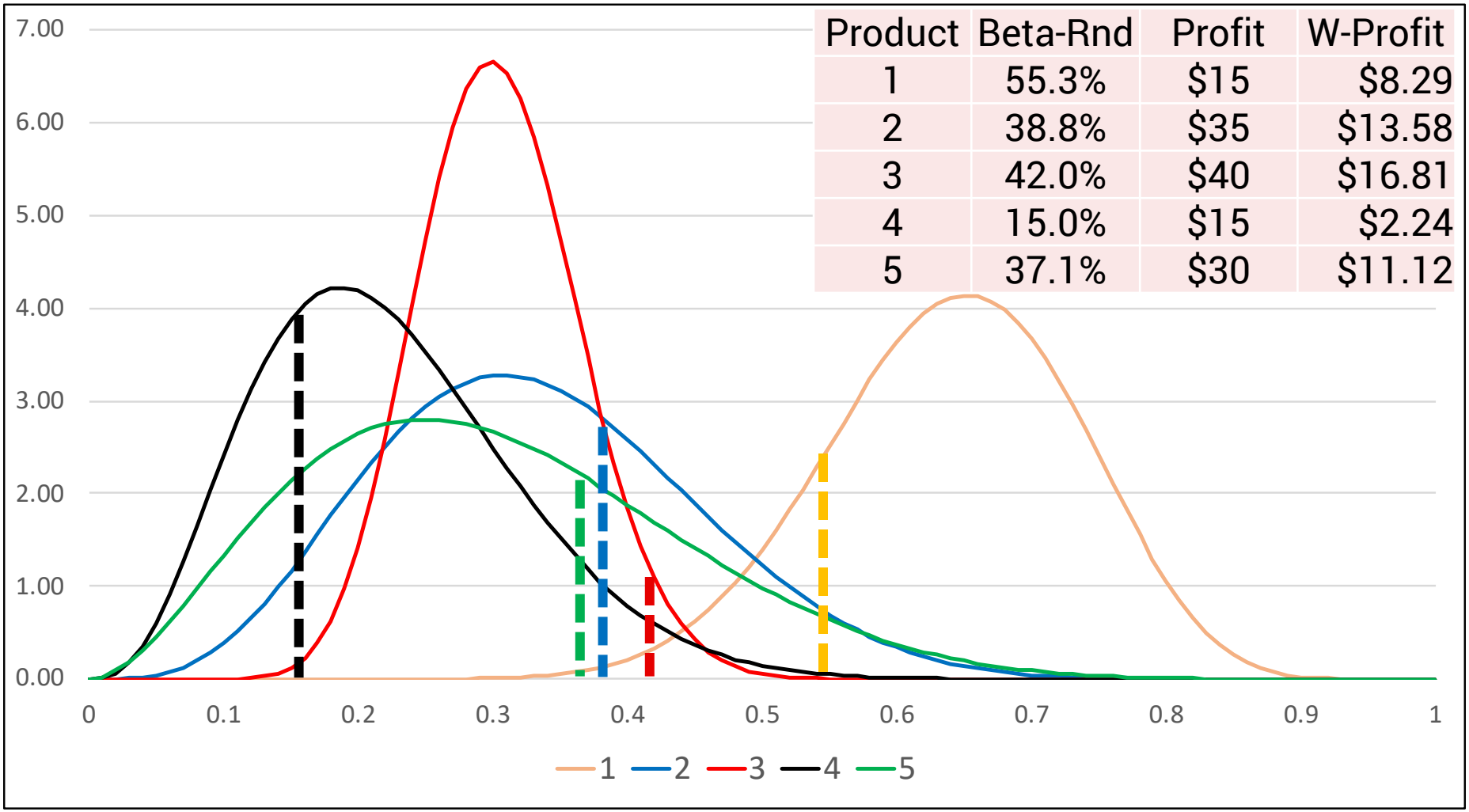
Display selection – Random Draw



Product	CR	Beta-Rnd
1	64.0%	55.3%
2	33.3%	38.8%
3	30.5%	42.0%
4	22.2%	15.0%
5	30.0%	37.1%

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

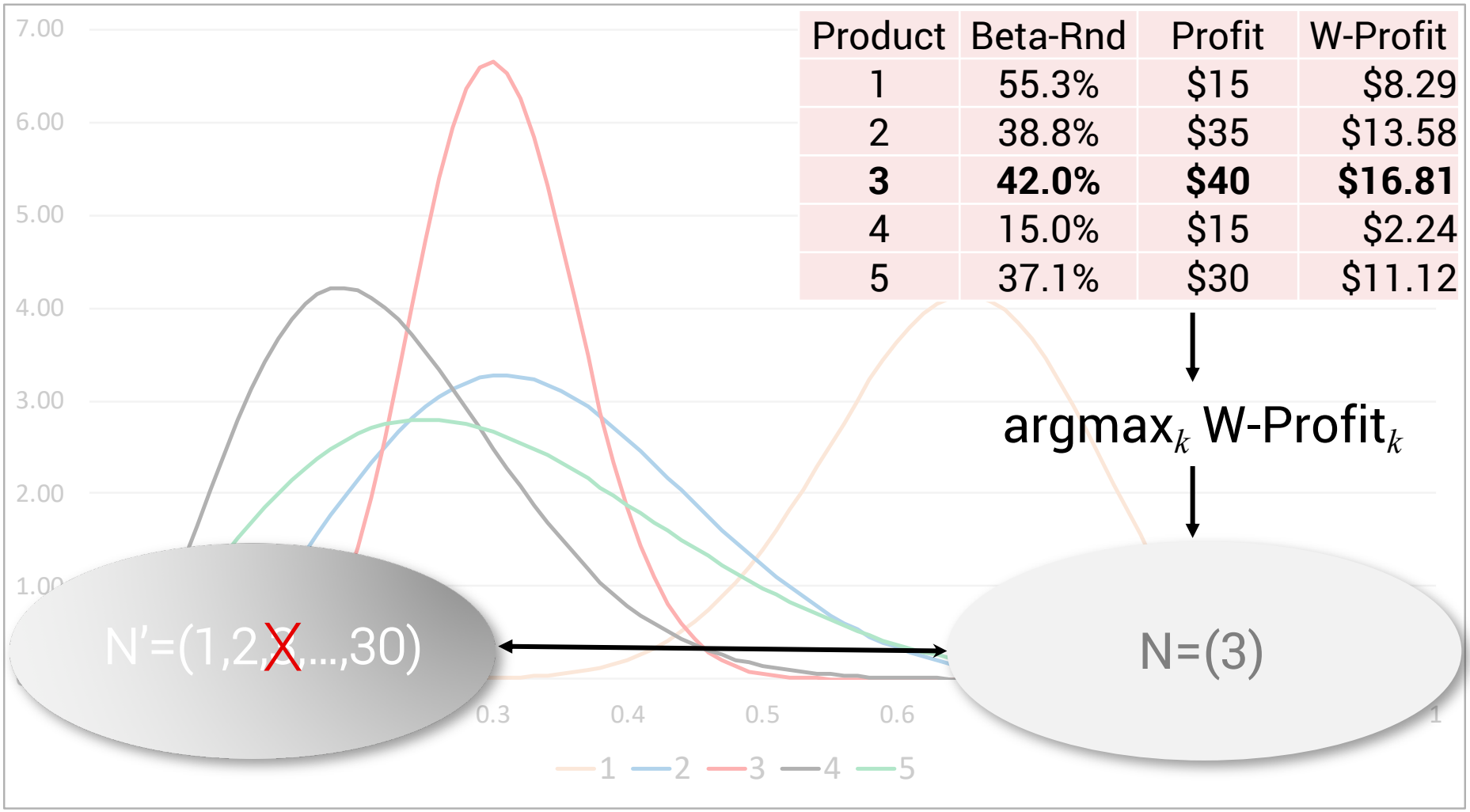
Display selection – Weighted by profit



Product	Beta-Rnd	Profit	W-Profit
1	55.3%	$15	$8.29
2	38.8%	$35	$13.58
3	42.0%	$40	$16.81
4	15.0%	$15	$2.24
5	37.1%	$30	$11.12

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Display selection – Winner

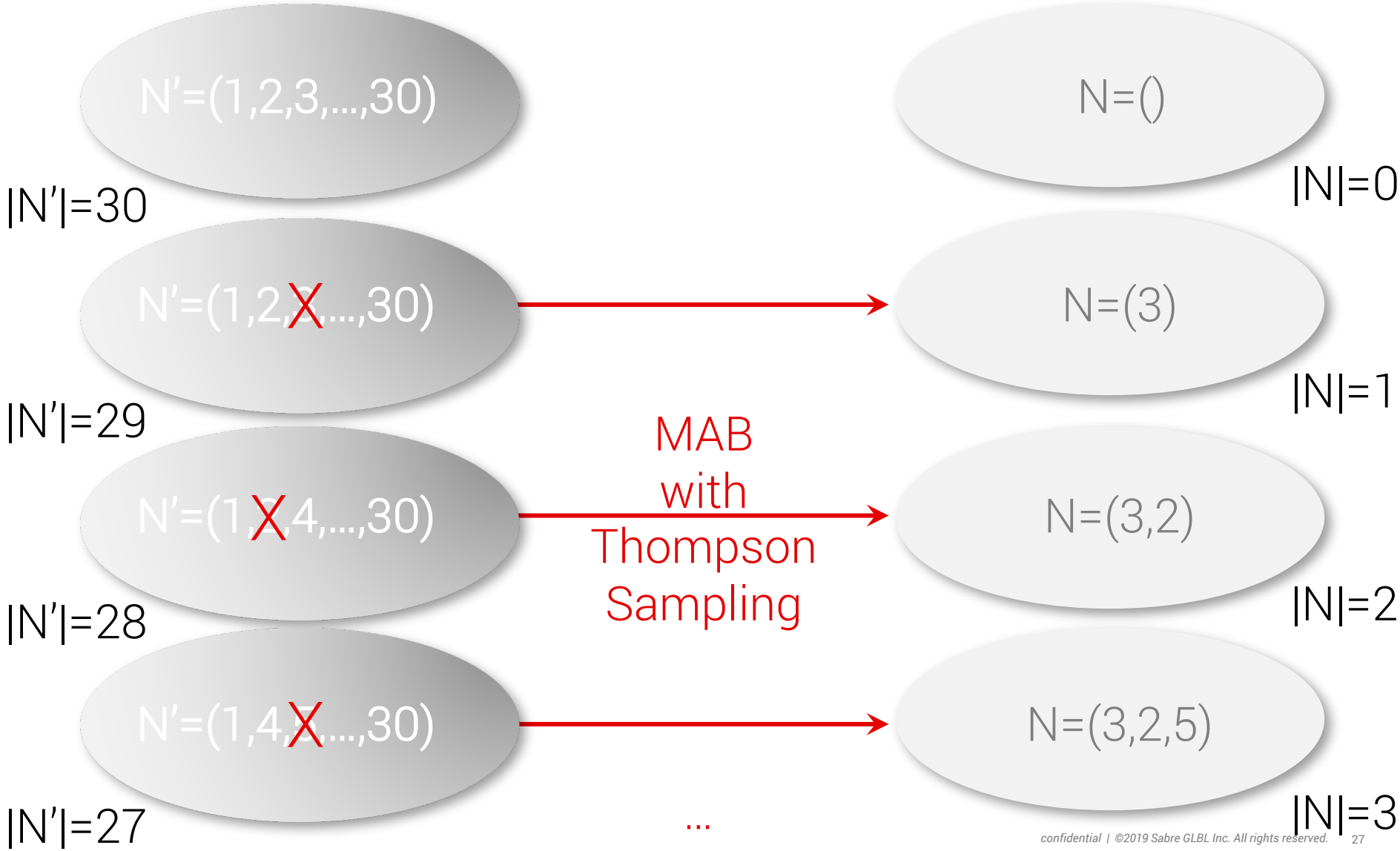


Product	Beta-Rnd	Profit	W-Profit
1	55.3%	$15	$8.29
2	38.8%	$35	$13.58
3	**42.0%**	**$40**	**$16.81**
4	15.0%	$15	$2.24
5	37.1%	$30	$11.12

argmax_k W-Profit$_k$

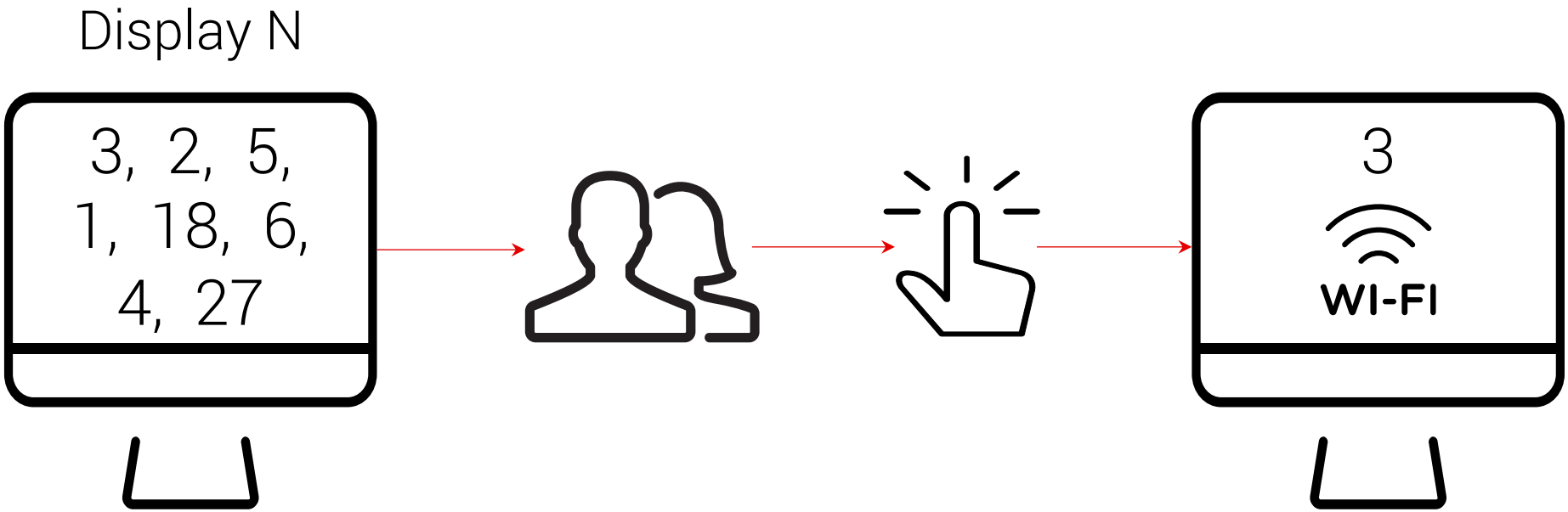
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Display selection without replacement



confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Real life experimenting



Update posteriors

$\alpha_3 = \alpha_3 + 1$
$\mathcal{B}_3 = \mathcal{B}_3$

$\alpha_k = \alpha_k$
$\mathcal{B}_k = \mathcal{B}_k + 1$
k in N-{3}

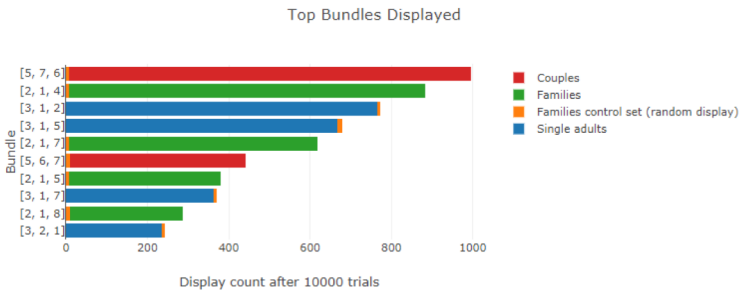
Select products without replacement based on Wrnd

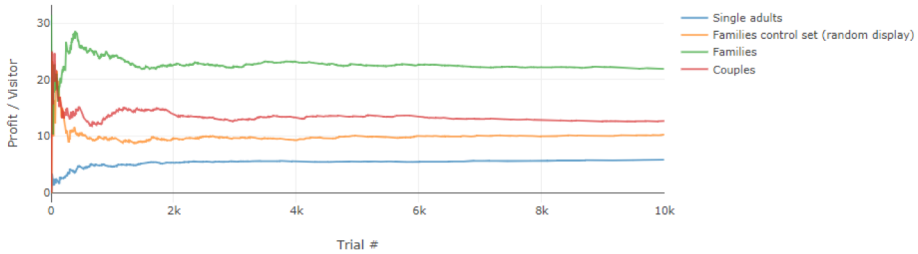
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

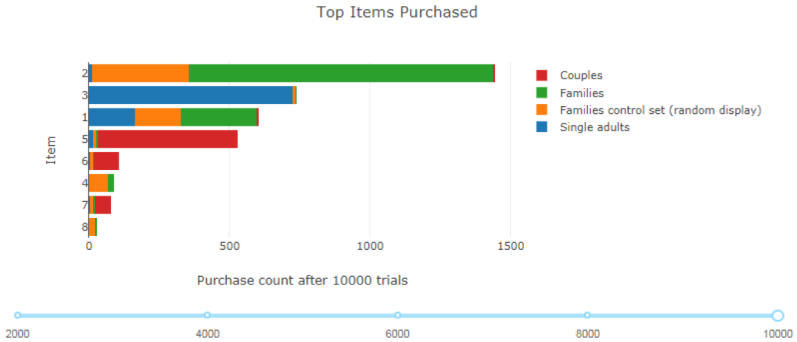
Simulation studies with segmentation

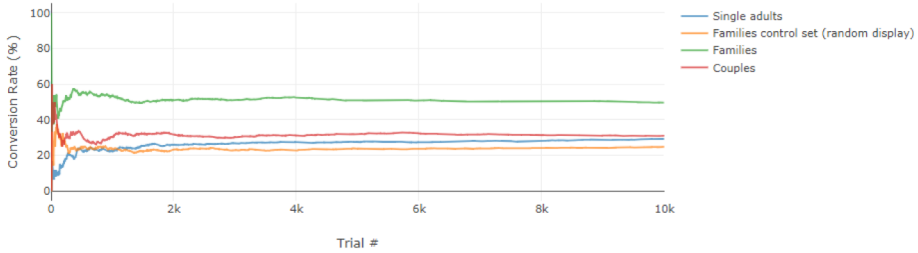
Multi-Armed Bandit

Hotel Website Display Problem









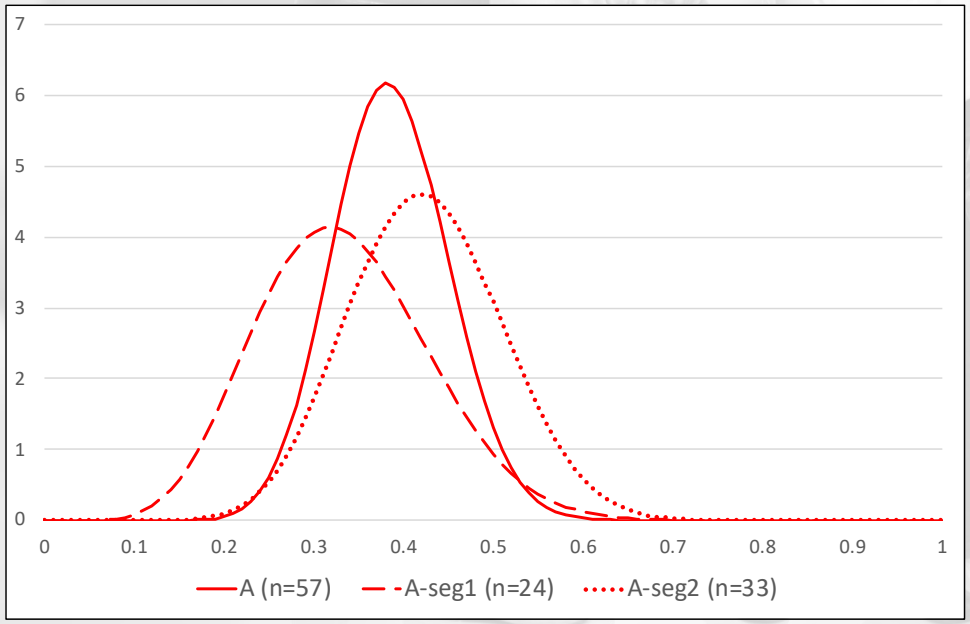
confidential | ©2019 Sabre GLBL Inc. All rights reserved.

NEXT STEPS

How can we improve?

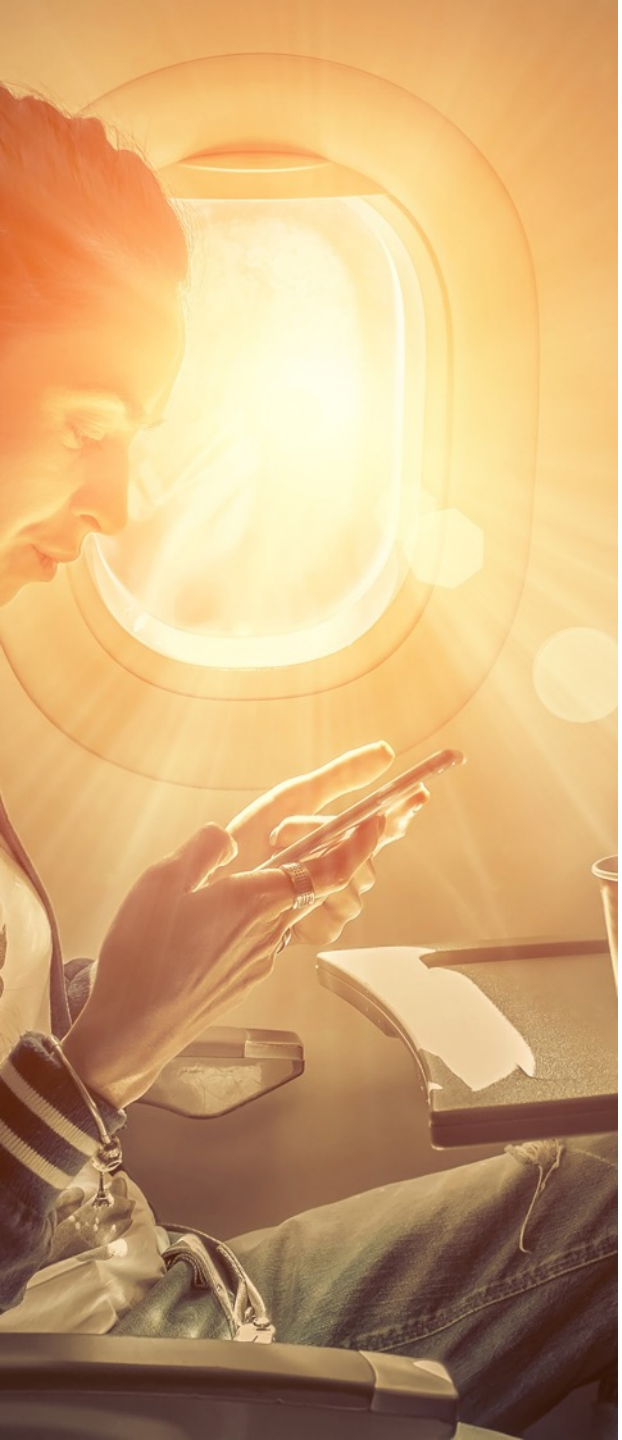
Segmentation

Pooling may hide actual patterns



- Discover what is working best for each customer
- Major difficulties:
  - Revealed personal data
  - Privacy
- Trip purpose segmentation:
  - Accounts for different profiles
  - Reveals actual purpose

confidential | ©2019 Sabre GLBL Inc. All rights reserved.



Contextual Bandit

- **Context:** A customer comes to our webpage (we know the IP address, history of previous visits, more info from his membership account if exists, etc.)
- **Action:** We select what content to present (from among available products)
- **Feedback:** The customer reacts (clicks, not clicks, etc.)
- Develop a strategy to action given context

confidential | ©2019 Sabre GLBL Inc. All rights reserved.

Assortment selection

- Products are dependent
- Assortment optimization with exploration & exploitation
- Customer reaction on a product is influenced by other products in the display

SUMMARY

The most competitive times for travel industry

Next generation retailing

- **Deliver** memorable and personalized guest experiences
- **Target** and **engage** guests thoughtfully at different stages of their travel journey
- **Streamline** targeting, offer creation, pricing and content management
- **Offer** them the most relevant products and services to meet their unique needs

Drive true brand differentiation

Q&A

Hunkar Toyoglu
Sr.Director
Advanced Analytics Services
hunkar.toyoglu@sabre.com

Richard Ratliff
Sr.Research Scientist
Sabre Labs and Research
richard.ratliff@sabre.com