On the Complexity of Differentially Private Best-Arm Identification with Fixed Confidence

Achraf Azize, **Marc Jourdan**, **Aymen Al Marjani** and **Debabrota Basu**

Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

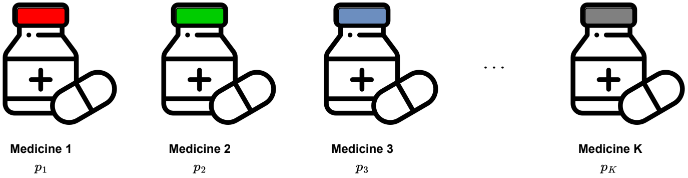
Outline

1. A short tour of the Best-Arm Identification (BAI) setting
2. Defining Privacy for BAI
3. Quantifying the Hardness of DP-BAI
4. Near-Optimal Algorithm for DP-BAI
5. Conclusion and Future Work

A short tour of BAI

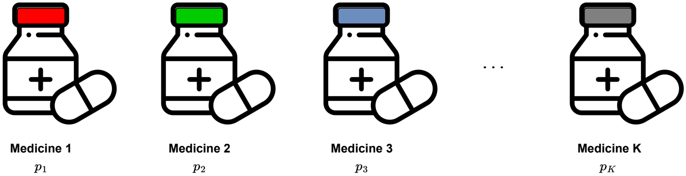
Sequential Decision Making

under Uncertainty: Multi-armed Bandits [Thompson, 1933]



Sequential Decision Making

under Uncertainty: Multi-armed Bandits [Thompson, 1933]

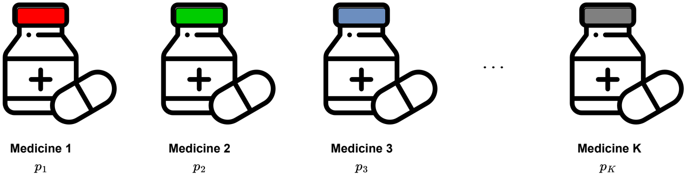


For the t-th patient in the study:

1. The doctor π chooses a Medicine $a_t \in \{1, \ldots, K\}$
2. The doctor observes a reward $r_t \in \{0, 1\}$ such that $r_t \sim \text{Bernouli}(p_{a_t})$

Sequential Decision Making

under Uncertainty: Multi-armed Bandits [Thompson, 1933]



For the t-th patient in the study:

1. The doctor π chooses a Medicine $a_t \in \{1, \ldots, K\}$
2. The doctor observes a reward $r_t \in \{0, 1\}$ such that $r_t \sim \text{Bernouli}(p_{a_t})$

Objective: Identify the medicine with the highest mean $a^\star \triangleq \text{argmax}_{a \in [K]} p_a$

Performance Measure for BAI

δ-correctness and Stopping Time

Goal: (a) Stop the interaction at time τ
(b) Recommend an arm $\widehat{a} \in [K]$

Performance Measure for BAI

δ-correctness and Stopping Time

Goal: (a) Stop the interaction at time τ
(b) Recommend an arm $\widehat{a} \in [K]$

Definition: A BAI strategy π is δ-correct for a class of instances $\mathcal{M}$, if

$$\mathbb{P}_{\boldsymbol{\nu},\pi}(\tau < \infty, \widehat{a} = a^\star(\boldsymbol{\nu})) \geq 1 - \delta$$

for every environment $\boldsymbol{\nu} = \{p_1, \ldots, p_K\} \in \mathcal{M}$.

Hardness of BAI

Theorem: [Garivier and Kaufmann, 2016] For any δ-correct BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq T^\star_{\mathrm{KL}}(\boldsymbol{\nu}) \log(1/3\delta),$$

and $T^\star_{\mathrm{KL}}(\boldsymbol{\nu}) \triangleq \left(\sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} \omega_a \mathrm{KL}(\nu_a, \lambda_a)\right)^{-1}$

Hardness of BAI

Theorem: [Garivier and Kaufmann, 2016] For any δ-correct BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq T^{\star}_{\mathrm{KL}}(\boldsymbol{\nu}) \log(1/3\delta),$$

and $T^{\star}_{\mathrm{KL}}(\boldsymbol{\nu}) \triangleq \left(\sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} \omega_a \mathrm{KL}(\nu_a, \lambda_a)\right)^{-1} \approx \sum_a \frac{1}{(\mu_{a^\star} - \mu_a)^2}$

Hardness of BAI

Theorem: [Garivier and Kaufmann, 2016] For any δ-correct BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq T^{\star}_{\mathrm{KL}}(\boldsymbol{\nu}) \log(1/3\delta),$$

and $T^{\star}_{\mathrm{KL}}(\boldsymbol{\nu}) \triangleq \left(\sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} \omega_a \mathrm{KL}(\nu_a, \lambda_a)\right)^{-1} \approx \sum_a \frac{1}{(\mu_{a^\star} - \mu_a)^2}$

Theorem: There exists an algorithm π such that

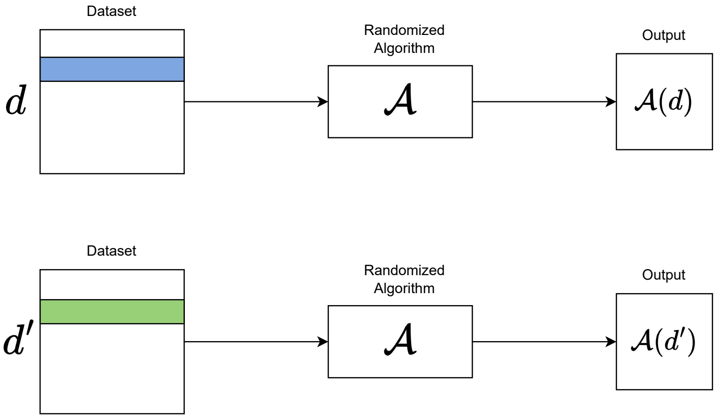
$$\lim_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau]}{\log(1/\delta)} = T^{\star}_{\mathrm{KL}}(\boldsymbol{\nu})$$

Example of such algorithms: Track And Stop [Garivier and Kaufmann, 2016], DKM [Degenne et al., 2019], **Top Two Algorithm [Jourdan et al., 2022]**.

Defining Privacy for BAI

Differential Privacy

Intuition: Indistinguishability from the mass



Differential Privacy

Intuition: Indistinguishability from the mass

Definition: [Dwork and Roth, 2014] A randomised algorithm $\mathcal{A}$ satisfies ϵ-DP if for any two neighbouring datasets d and d' that differ only in one row, i.e $d \sim d'$, and for all sets of output $\mathcal{O} \subseteq \mathrm{Range}(\mathcal{A})$,

$$\Pr[\mathcal{A}(d) \in \mathcal{O}] \leq e^{\epsilon} \Pr\left[\mathcal{A}\left(d'\right) \in \mathcal{O}\right]$$

Differential Privacy

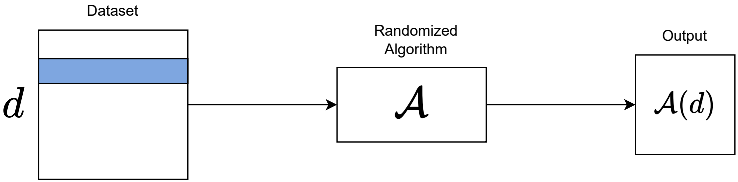
Intuition: Indistinguishability from the mass

Definition: [Dwork and Roth, 2014] A randomised algorithm $\mathcal{A}$ satisfies ϵ-DP if for any two neighbouring datasets d and d' that differ only in one row, i.e $d \sim d'$, and for all sets of output $\mathcal{O} \subseteq \mathrm{Range}(\mathcal{A})$,

$$\Pr[\mathcal{A}(d) \in \mathcal{O}] \leq e^{\epsilon} \Pr\left[\mathcal{A}\left(d'\right) \in \mathcal{O}\right]$$

Privacy in BAI: Rewards may contain sensitive information about individuals. *A patient's reaction to a medicine can reveal sensitive information about their health conditions.*

BAI with DP



Ingredients to specify:

- The randomized algorithm
- The private input dataset
- The output

BAI with DP

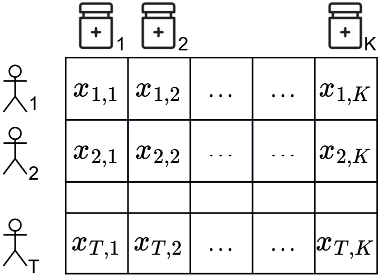
A BAI strategy π consists of:

- A pair of sampling and stopping rules $(\mathrm{S}_t)_{t\geq 1}$:
  - For $a \in [K]$, $\mathrm{S}_t\,(a \mid \mathcal{H}_{t-1})$ is the probability of playing action a given the history $\mathcal{H}_{t-1}$
  - $\mathrm{S}_t\,(\top \mid \mathcal{H}_{t-1})$ is the probability of the algorithm halting given $\mathcal{H}_{t-1}$

- A recommendation rule $(\mathrm{Rec}_t)_{t>1}$:
  - For $a \in [K]$, $\mathrm{Rec}_t\,(a \mid \mathcal{H}_{t-1})$ is the probability of returning action a as a guess for the best action given $\mathcal{H}_{t-1}$.

BAI with DP

The private dataset $\underline{\mathbf{d}}^T$ is



BAI with DP

The private dataset $\underline{\mathbf{d}}^T$ is

	1	2			K
1	$x_{1,1}$	$x_{1,2}$	$\dots$	$\dots$	$x_{1,K}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$\cdots$	$x_{2,K}$
T	$x_{T,1}$	$x_{T,2}$	$\dots$	$\dots$	$x_{T,K}$

When a_t is recommended to Patient p_t, only $r_t \triangleq x_{t,a_t}$ is observed

BAI with DP

The private dataset $\underline{\mathbf{d}}^T$ is

	1	2			K
1	$x_{1,1}$	$x_{1,2}$	$\dots$	$\dots$	$x_{1,K}$
2	$x_{2,1}$	$x_{2,2}$	$\cdots$	$\cdots$	$x_{2,K}$
T	$x_{T,1}$	$x_{T,2}$	$\dots$	$\dots$	$x_{T,K}$

When a_t is recommended to Patient p_t, only $r_t \triangleq x_{t,a_t}$ is observed

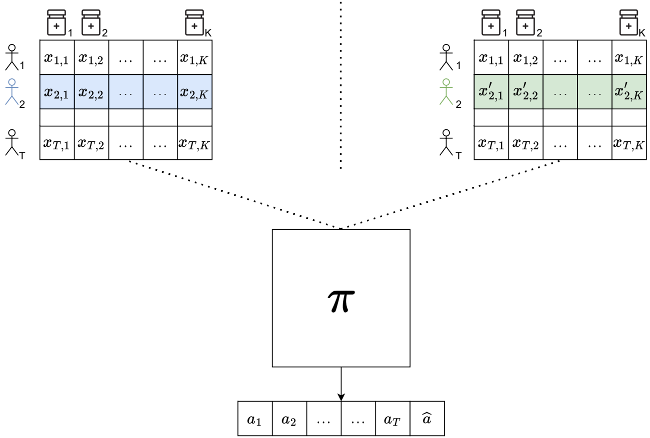
Finally, the mechanism induced by the interaction is

$$\pi(\underline{a}^T, \widehat{a}, T \mid \underline{\mathbf{d}}^T) \triangleq \mathrm{Rec}_{T+1}(\widehat{a} \mid \mathcal{H}_T)\, \mathrm{S}_{T+1}(\top \mid \mathcal{H}_T) \prod_{t=1}^{T} \mathrm{S}_t(a_t \mid \mathcal{H}_{t-1})$$

ϵ-global DP BAI

Definition: π satisfies ϵ-**global DP**, if $\forall T \geq 1$, $\forall \underline{\mathbf{d}}^T \sim \underline{\mathbf{d}'}^T, \forall \underline{a}^T$ and $\widehat{a}$,

$$\pi(\underline{a}^T, \widehat{a}, T \mid \underline{\mathbf{d}}^T) \leq e^{\epsilon} \pi(\underline{a}^T, \widehat{a}, T \mid \underline{\mathbf{d}'}^T).$$



Main Question and Contributions

Main Question: What is the cost of ϵ-global DP in BAI?

Contributions:

- We provide a lower bound on the sample complexity of any δ-correct ϵ-global DP BAI strategy
- We design a near-optimal algorithm matching the sample complexity lower bound, up to multiplicative constants

Quantifying the Hardness of DP-BAI

Lower Bound

Theorem: For any δ-correct ϵ-global DP BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq \max\left(T^\star_{\mathrm{KL}}(\boldsymbol{\nu}), \frac{1}{6\epsilon} T^\star_{\mathrm{TV}}(\boldsymbol{\nu})\right) \log(1/3\delta),$$

$\left(T^\star_{\mathrm{d}}(\boldsymbol{\nu})\right)^{-1} \triangleq \sup_{\omega \in \boldsymbol{\Sigma}_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} \omega_a \mathrm{d}(\nu_a, \lambda_a)$, d is either KL or TV.

Lower Bound

Our Results

Theorem: For any δ-correct ϵ-global DP BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq \max\left(T^\star_{\mathrm{KL}}(\boldsymbol{\nu}), \frac{1}{6\epsilon} T^\star_{\mathrm{TV}}(\boldsymbol{\nu})\right) \log(1/3\delta),$$

$\left(T^\star_{\mathrm{d}}(\boldsymbol{\nu})\right)^{-1} \triangleq \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} \omega_a \mathrm{d}(\nu_a, \lambda_a)$, d is either KL or TV.

$$T^\star_{\mathrm{KL}}(\boldsymbol{\nu}) \approx \sum_a \frac{1}{(\mu_{a^\star} - \mu_a)^2} \quad \text{and} \quad T^\star_{\mathrm{TV}}(\boldsymbol{\nu}) \approx \sum_a \frac{1}{\mu_{a^\star} - \mu_a}$$

$$T^\star_{\mathrm{TV}}(\boldsymbol{\nu}) \geq \sqrt{2 T^\star_{\mathrm{KL}}(\boldsymbol{\nu})}$$

DP and Total Variation

Intuition: Stochastic Group Privacy

- d and d' differ in 1 sample $\rightarrow \exp(\epsilon)$

DP and Total Variation

Intuition: Stochastic Group Privacy

- d and d' differ in 1 sample $\rightarrow \exp(\epsilon)$
- d and d' differ in k samples $\rightarrow \exp(k\epsilon)$

DP and Total Variation

Intuition: Stochastic Group Privacy

- d and d' differ in 1 sample $\rightarrow \exp(\epsilon)$
- d and d' differ in k samples $\rightarrow \exp(k\epsilon)$
- Sample $d \sim \otimes^n P$ and $d' \sim \otimes^n Q \rightarrow \exp(nTV(P, Q)\epsilon)$

DP and Total Variation

Intuition: Stochastic Group Privacy

- d and d' differ in 1 sample $\rightarrow \exp(\epsilon)$
- d and d' differ in k samples $\rightarrow \exp(k\epsilon)$
- Sample $d \sim \otimes^n P$ and $d' \sim \otimes^n Q \rightarrow \exp(nTV(P, Q)\epsilon)$
- Sample $d \sim \otimes_{i=1}^n P_i$ and $d' \sim \otimes_{i=1}^n Q_i \rightarrow \exp\left(\sum_{i=1}^n TV(P_i, Q_i)\epsilon\right)$

Lower Bound

Discussion

$$\mathbb{E}_{\boldsymbol{\nu},\pi}[\tau] \geq \max\left(T^\star_{\mathrm{KL}}(\boldsymbol{\nu}), \frac{1}{6\epsilon} T^\star_{\mathrm{TV}}(\boldsymbol{\nu})\right) \log(1/3\delta)$$

Two hardness regimes depending on ϵ and the environment $\boldsymbol{\nu}$:

- *Low-privacy regime*: When $\epsilon > \frac{T^\star_{\mathrm{TV}}(\boldsymbol{\nu})}{6 T^\star_{\mathrm{KL}}(\boldsymbol{\nu})}$, the lower bound retrieves the non-private $T^\star_{\mathrm{KL}}(\boldsymbol{\nu})$ lower bound and **privacy can be achieved for free**.
- *High-privacy regime:* When $\epsilon < \frac{T^\star_{\mathrm{TV}}(\boldsymbol{\nu})}{6 T^\star_{\mathrm{KL}}(\boldsymbol{\nu})}$, the lower bound becomes $\frac{1}{6\epsilon} T^\star_{\mathrm{TV}}(\boldsymbol{\nu})$ and ϵ-global DP δ-BAI requires more samples than non-private ones.

Near-Optimal Algorithm for DP-BAI

Algorithm Design

Top Two Algorithm

Algorithm Design

Top Two Algorithm

The Top Two sampling rule:

- Choosing a **leader** $B_n \in [K]$
- Choosing a **challenger** $C_n \in [K] \setminus \{B_n\}$
- Sampling B_n with probability $\frac{1}{2}$, else sampling C_n

Algorithm Design

Top Two Algorithm

The Top Two sampling rule:

- Choosing a **leader** $B_n \in [K]$
- Choosing a **challenger** $C_n \in [K] \setminus \{B_n\}$
- Sampling B_n with probability $\frac{1}{2}$, else sampling C_n

Leader: Empirical Best; $B_n = \operatorname{argmax}_{a \in [K]} \widehat{\mu}_{n,a}$.

Algorithm Design

Top Two Algorithm

The Top Two sampling rule:

- Choosing a **leader** $B_n \in [K]$
- Choosing a **challenger** $C_n \in [K] \setminus \{B_n\}$
- Sampling B_n with probability $\frac{1}{2}$, else sampling C_n

Leader: Empirical Best; $B_n = \mathrm{argmax}_{a \in [K]} \widehat{\mu}_{n,a}$.

Challenger: Transportation Cost;

$$C_n = \operatorname*{argmin}_{j \neq B_n} W_n(B_n, j) \,.$$

Algorithm Design

Top Two Algorithm

The Top Two sampling rule:

- Choosing a **leader** $B_n \in [K]$
- Choosing a **challenger** $C_n \in [K] \setminus \{B_n\}$
- Sampling B_n with probability $\frac{1}{2}$, else sampling C_n

The recommendation rule:

$$\widehat{a}_n = \operatorname*{argmax}_{a \in [K]} \widehat{\mu}_{n,a}$$

Algorithm Design

Top Two Algorithm

The Top Two sampling rule:

- Choosing a **leader** $B_n \in [K]$
- Choosing a **challenger** $C_n \in [K] \setminus \{B_n\}$
- Sampling B_n with probability $\frac{1}{2}$, else sampling C_n

The recommendation rule:

$$\widehat{a}_n = \operatorname*{argmax}_{a \in [K]} \widehat{\mu}_{n,a}$$

The stopping rule is a GLR test

$$\tau_\delta = \inf\{n \mid \min_{j \neq \widehat{a}_n} W_n(\widehat{a}_n, j) > c(n, \delta)\} ,$$

Algorithm Design

Private Top Two

To make the Top Two algorithm satisfy ϵ-global DP, we

Algorithm Design

Private Top Two

To make the Top Two algorithm satisfy ϵ-global DP, we

- Estimate the sequence of empirical means $(\widehat{\mu}_{a,n})$ privately, i.e. $(\tilde{\mu}_{a,n}) = (\widehat{\mu}_{a,n}) + \frac{1}{\epsilon} Lap$, using
  - Per-arm doubling
  - Forgetting
  - Adding calibrated Laplace noise

Algorithm Design

Private Top Two

To make the Top Two algorithm satisfy ϵ-global DP, we

- Estimate the sequence of empirical means $(\widehat{\mu}_{a,n})$ privately, i.e. $(\tilde{\mu}_{a,n}) = (\widehat{\mu}_{a,n}) + \frac{1}{\epsilon} Lap$, using
  - Per-arm doubling
  - Forgetting
  - Adding calibrated Laplace noise

- Calibrate for the noise in the components:
  - The sampling rule: leader and challenger based on the private $(\tilde{\mu}_{a,n})$
  - The recommendation rule: Recommend $\widehat{a}_n = \operatorname{argmax}_{a \in [K]} \tilde{\mu}_{n,a}$
  - The stopping rule: re-calibrate the GLR threshold $\tilde{c}(n, \delta) = c(n, \delta) + \frac{1}{\epsilon} c_2(n, \delta)$

Algorithm Design

Privacy and sample complexity

Theorem: For Bernoulli instances verifying that $\exists C \geq 1$ such that $\Delta_{\max}/\Delta_{\min} \leq C$, AdaP-TT is ϵ-global DP, δ-correct and satisfies

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\mu}[\tau_\delta]}{\log(1/\delta)} \leq c \ \max\left\{ T^\star_{\mathrm{KL}}(\boldsymbol{\mu}), C \frac{T^\star_{\mathrm{TV}}(\boldsymbol{\mu})}{\epsilon} \right\} .$$

where c is a universal constant.

☞ Matches the lower bound up to constants

Experimental Analysis

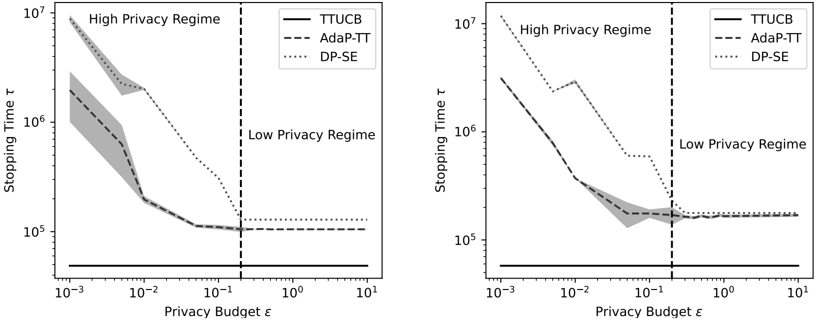


Figure: Evolution of the stopping time τ of AdaP-TT, DP-SE, and TTUCB with respect to the privacy budget ϵ for $\delta = 10^{-2}$ on two Bernoulli instances. The shaded vertical line separates the two privacy regimes. AdaP-TT outperforms DP-SE.

Conclusion and Future Work

Conclusion and Future Work

Conclusion: We derive sample complexity lower bounds and matching upper bounds for BAI with ϵ-global DP.

Future Work:

- Close the multiplicative gap between the lower and upper bounds.
- Extend the analysis to other DP settings, like (ϵ, δ)-DP and Rényi-DP.
- Extend the analysis to other trust models, like local DP and shuffle DP.

Thank you for your time

Questions!

Bibliography I

Degenne, R., Koolen, W. M., and Ménard, P. (2019). Non-asymptotic pure exploration by solving games.
Advances in Neural Information Processing Systems, 32.

Dwork, C. and Roth, A. (2014).
The algorithmic foundations of differential privacy.
Foundations and Trends® in Theoretical Computer Science, 9(3–4):211–407.

Garivier, A. and Kaufmann, E. (2016).
Optimal best arm identification with fixed confidence.
In Conference on Learning Theory, pages 998–1027. PMLR.

Jourdan, M., Degenne, R., Baudry, D., de Heide, R., and Kaufmann, E. (2022).
Top two algorithms revisited.
Advances in Neural Information Processing Systems, 35:26791–26803.

Thompson, W. R. (1933).
On the likelihood that one unknown probability exceeds another in view of the evidence of two samples.
Biometrika, 25(3-4):285–294.