# New tools from the bandit literature to improve A/B Testing

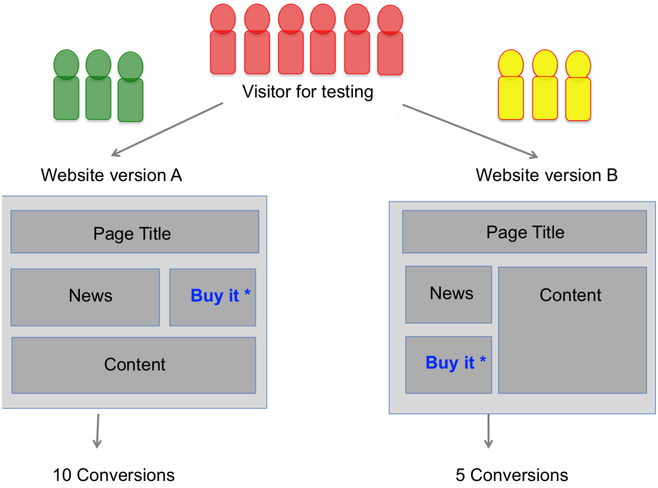
Emilie Kaufmann,

joint work with Aurélien Garivier & Olivier Cappé

Centre de Recherche en Informatique, Signal et Automatique de Lille

RecSys Meetup,
March 23rd, 2016

# Motivation



Visitor for testing
Website version A
Page Title
News
Buy it *
Content
10 Conversions
Website version B
Page Title
News
Content
Buy it *
5 Conversions

# A/B Testing

A way to do A/B Testing:

- allocate $n_A$ users to page A and $n_B$ users to page B
- perform a statistical test of "A better than B"

# A/B Testing

A way to do A/B Testing:

- allocate $n_A$ users to page A and $n_B$ users to page B
- perform a statistical test of "A better than B"

A variant: fully adaptive A/B Testing

- sequentially choose which version to allocate to each visitor
- adaptively choose when to stop the experiment

➜ multi-armed bandit model

# A/B/C... testing as a Best Arm Identification problem

$K$ arms = $K$ probability distributions ($\nu_a$ has mean $\mu_a$)


$\nu_1$ $\nu_2$ $\nu_3$ $\nu_4$ $\nu_5$

$$a^* = \underset{a=1,\dots,K}{\text{argmax}}\ \mu_a$$

For the $t$-th user,

- allocate a version (arm) $A_t \in \{1, \dots, K\}$
- observe a feedback $X_t \sim \nu_{A_t}$

Goal: design

- a sequential sampling rule: $A_{t+1} = F_t(A_1, X_1, \dots, A_t, X_t)$,
- a stopping rule $\tau$
- a recommendation rule $\hat{a}_\tau$

such that $\mathbb{P}(\hat{a}_\tau = a^*) \geq 1 - \delta$ and $\tau$ is as small as possible.

# Outline

1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# Outline

1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# PAC algorithms in one-parameter bandit models

$\mathcal{P} = \{\nu_\mu, \mu \in \mathcal{I}\}$ set of distributions parametrized by their mean

**Example:** Bernoulli, Poisson, Gaussian (known variance)

$$\nu_{\mu_1}, \dots, \nu_{\mu_K} \in \mathcal{P}^K \quad \Leftrightarrow \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$$

$$\mathcal{S} = \left\{ \boldsymbol{\mu} \in \mathcal{I}^K : \exists a \in \{1, \dots, K\} : \mu_a > \max_{i \neq a} \mu_i \right\}$$

- A strategy is $\delta$-PAC (on $\mathcal{S}$) if

$$\forall \nu \in \mathcal{S}, \quad \mathbb{P}_\nu(\hat{a}_\tau = a^*) \geq 1 - \delta.$$

# PAC algorithms in one-parameter bandit models

$\mathcal{P} = \{\nu_\mu, \mu \in \mathcal{I}\}$ set of distributions parametrized by their mean

**Example:** Bernoulli, Poisson, Gaussian (known variance)

$$\nu_{\mu_1}, \dots, \nu_{\mu_K} \in \mathcal{P}^K \quad \Leftrightarrow \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$$

$$\mathcal{S} = \left\{ \boldsymbol{\mu} \in \mathcal{I}^K : \exists a \in \{1, \dots, K\} : \mu_a > \max_{i \neq a} \mu_i \right\}$$

- A strategy is $\delta$-PAC (on $\mathcal{S}$) if

$$\forall \nu \in \mathcal{S}, \quad \mathbb{P}_\nu(\hat{a}_\tau = a^*) \geq 1 - \delta.$$

➜ What is the optimal sample complexity of a $\delta$-PAC strategy?

$$\inf_{\delta\text{-PAC}} \mathbb{E}_{\boldsymbol{\mu}}[\tau]?$$

# The optimal sample complexity

To answer this question, we need

- ➜ a lower bound on $\mathbb{E}_\nu[\tau]$ for any $\delta$-PAC strategy
- ➜ a $\delta$-PAC strategy such that $\mathbb{E}_\nu[\tau]$ matches this bound

**State-of-the-art:** $\delta$-PAC algorithms for which

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] = O\left(H(\boldsymbol{\mu})\log\frac{1}{\delta}\right), \quad H(\boldsymbol{\mu}) = \frac{1}{(\mu_2-\mu_1)^2} + \sum_{a=2}^{K}\frac{1}{(\mu_a-\mu_1)^2}$$

[Even Dar et al. 2006, Kalyanakrishnan et al. 2012]

- ➜ the optimal sample complexity is not identified...

**Notation: Kullback-Leibler divergence**

$$d(\mu,\mu') := \mathrm{KL}(\nu^{\mu},\nu^{\mu'}) = \mathbb{E}_{X\sim\nu^{\mu}}\left[\log\frac{d\nu^{\mu}}{d\nu^{\mu'}}(X)\right]$$

is the KL-divergence between the distributions of mean $\mu$ and $\mu'$.

# Outline

1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# Lower bound

- A first (easy to interpret) result

**Theorem** [Kaufmann, Cappé, Garivier 2015]

For any $\delta$-PAC algorithm,

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \geq \left( \frac{1}{d(\mu_1, \mu_2)} + \sum_{a=2}^{K} \frac{1}{d(\mu_a, \mu_1)} \right) \log \left( \frac{1}{2.4\delta} \right)$$

# Lower bound

- A first (easy to interpret) result

**Theorem** [Kaufmann, Cappé, Garivier 2015]

For any $\delta$-PAC algorithm,

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \geq \left( \frac{1}{d(\mu_1, \mu_2)} + \sum_{a=2}^{K} \frac{1}{d(\mu_a, \mu_1)} \right) \log \left( \frac{1}{2.4\delta} \right)$$

- A tighter (non explicit) lower bound

**Theorem** [Kaufmann and Garivier, 2016]

$\mathrm{Alt}(\boldsymbol{\mu}) := \{\boldsymbol{\lambda} : a^*(\boldsymbol{\lambda}) \neq a^*(\boldsymbol{\mu})\}$. For any $\delta$-PAC algorithm,

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \geq T^*(\boldsymbol{\mu}) \log \left( \frac{1}{2.4\delta} \right),$$

where

$$T^*(\boldsymbol{\mu})^{-1} = \sup_{w \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \left( \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right).$$

# A vector of optimal proportions

$$w^*(\boldsymbol{\mu}) := \underset{w \in \Sigma_K}{\text{argmax}} \inf_{\lambda \in \text{Alt}(\boldsymbol{\mu})} \left( \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a) \right)$$

is unique and represents the optimal proportions of draws: a strategy matching the lower bound should satisfy

$$\forall a \in \{1, \dots, K\}, \ \frac{\mathbb{E}_{\boldsymbol{\mu}}[N_a(\tau)]}{\mathbb{E}_{\boldsymbol{\mu}}[\tau]} = w_a^*(\boldsymbol{\mu}).$$

$N_a(t)$ : number of draws of arm $a$ up to time $t$

➜ we propose an efficient algorithm to compute $w^*(\boldsymbol{\mu})$

# Outline


1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# Sampling rule: Tracking the optimal proportions

$\hat{\boldsymbol{\mu}}(t) = (\hat{\mu}_1(t), \dots, \hat{\mu}_K(t))$: vector of empirical means

- Introducing

$$U_t = \{a : N_a(t) < \sqrt{t}\},$$

the arm sampled at round $t+1$ is

$$A_{t+1} \in \begin{cases} \underset{a \in U_t}{\text{argmin}}\ N_a(t) \text{ if } U_t \neq \emptyset & (\textit{forced exploration}) \\ \underset{1 \leq a \leq K}{\text{argmax}}\, [t\ w_a^*(\hat{\boldsymbol{\mu}}(t)) - N_a(t)] & (\textit{tracking}) \end{cases}$$

**Lemma**

Under the Tracking sampling rule,

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\lim_{t \to \infty} \frac{N_a(t)}{t} = w_a^*(\boldsymbol{\mu})\right) = 1.$$

# Stopping rule: performing statistical tests

High values of the Generalized Likelihood Ratio

$$Z_{a,b}(t) := \log \frac{\max_{\{\boldsymbol{\lambda}:\lambda_a \geq \lambda_b\}} \ell(X_1,\dots,X_t;\boldsymbol{\lambda})}{\max_{\{\boldsymbol{\lambda}:\lambda_a \leq \lambda_b\}} \ell(X_1,\dots,X_t;\boldsymbol{\lambda})},$$

reject the hypothesis that $(\mu_a < \mu_b)$.

We stop when one arm is accessed to be significantly larger than all other arms, according to a GLR Test:

$$\begin{aligned} \tau_\delta &= \inf\{t \in \mathbb{N} : \exists a \in \{1,\dots,K\}, \forall b \neq a, Z_{a,b}(t) > \beta(t,\delta)\} \\ &= \inf\left\{t \in \mathbb{N} : \max_{a \in \{1,\dots,K\}} \min_{b \neq a} Z_{a,b}(t) > \beta(t,\delta)\right\} \end{aligned}$$

Chernoff stopping rule [Chernoff 59]

# Stopping rule: an alternative interpretation

One has $Z_{a,b}(t) = -Z_{b,a}(t)$ and, if $\hat{\mu}_a(t) \geq \hat{\mu}_b(t)$,

$$Z_{a,b}(t) = N_a(t)\, d\left(\hat{\mu}_a(t), \hat{\mu}_{a,b}(t)\right) + N_b(t)\, d\left(\hat{\mu}_b(t), \hat{\mu}_{a,b}(t)\right),$$

where $\hat{\mu}_{a,b}(t) := \frac{N_a(t)}{N_a(t)+N_b(t)}\hat{\mu}_a(t) + \frac{N_b(t)}{N_a(t)+N_b(t)}\hat{\mu}_b(t)$.

## A link with the lower bound

$$\begin{aligned} \max_a \min_{b \neq a} Z_{a,b}(t) &= t \times \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\hat{\mu}(t))} \sum_{a=1}^{K} \frac{N_a(t)}{t} d(\hat{\mu}_a(t), \lambda_a) \\ &\simeq \frac{t}{T^*(\boldsymbol{\mu})} \end{aligned}$$

under a "good" sampling strategy (for $t$ large)

# An asymptotically optimal algorithm

**Theorem**

The Track-and-Stop strategy, that uses

- the Tracking sampling rule
- the Chernoff stopping rule with $\beta(t,\delta) = \log\left(\frac{2(K-1)t}{\delta}\right)$
- and recommends $\hat{a}_\tau = \underset{a=1\dots K}{\operatorname{argmax}}\ \hat{\mu}_a(\tau)$

is $\delta$-PAC for every $\delta \in ]0,1[$ and satisfies

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} = T^*(\boldsymbol{\mu}).$$

# Outline


1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

Two arms, $\mathcal{B}(\mu_1)$ and $\mathcal{B}(\mu_2)$

$$\mathbb{E}_\mu[\tau] \geq T^*(\boldsymbol{\mu}) \log\left(\frac{1}{2.4\delta}\right),$$

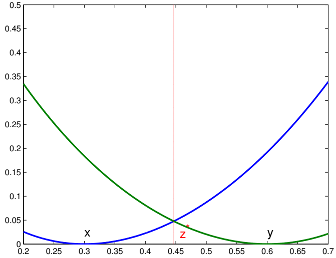
with

$$\begin{aligned} T^*(\boldsymbol{\mu})^{-1} &= \sup_{\alpha\in[0,1]} [\alpha d(\mu_1, \alpha\mu_1 + (1-\alpha)\mu_2) + \\ &\qquad (1-\alpha) d(\mu_2, \alpha\mu_1 + (1-\alpha)\mu_2)] \\ &= d_*(\mu_1, \mu_2), \end{aligned}$$

$$d_*(\mu_1, \mu_2) = d(\mu_1, z^*)$$

with $z^*$ defined by

$$d(\mu_1, z^*) = d(\mu_2, z^*)$$



# Algorithms

**Track-and-Stop**

- Sampling rule:

$$A_{t+1} = \underset{a=1,2}{\text{argmax}}\; d\left(\hat{\mu}_a(t), \frac{N_1(t)\hat{\mu}_1(t) + N_2(t)\hat{\mu}_2(t)}{N_1(t)+N_2(t)}\right)$$

- Stopping rule: stop after $t$ samples if

$$\sum_{a=1,2} N_a(t) d\left(\hat{\mu}_a(t), \frac{N_1(t)\hat{\mu}_1(t) + N_2(t)\hat{\mu}_2(t)}{N_1(t)+N_2(t)}\right) > \beta(t,\delta)$$

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \simeq \frac{1}{d_*(\mu_1,\mu_2)} \log\left(\frac{1}{\delta}\right)$$

# Algorithms

**Uniform sampling (and optimal stopping)**

- Sampling rule:

$$A_{t+1} = t \ [2]$$

- Stopping rule: stop after $t$ samples if

$$\sum_{a=1,2} N_a(t) d\left(\hat{\mu}_a(t), \frac{N_1(t)\hat{\mu}_1(t) + N_2(t)\hat{\mu}_2(t)}{N_1(t)+N_2(t)}\right) > \beta(t,\delta)$$

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \simeq \frac{1}{I_*(\mu_1,\mu_2)} \log\left(\frac{1}{\delta}\right)$$

with

$$I_*(\mu_1,\mu_2) = \frac{d\left(\mu_1, \frac{\mu_1+\mu_2}{2}\right) + d\left(\mu_1, \frac{\mu_1+\mu_2}{2}\right)}{2}.$$

**Remark:** $I_*(\mu_1,\mu_2)$ very close to $d_*(\mu_1,\mu_2)$

➜ uniform sampling is close to optimal

# Outline


1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# An optimal algorithm

Two arms, $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ $\sigma_1, \sigma_2$ known

$$\mathbb{E}_\mu[\tau] \geq \frac{2(\sigma_1^2 + \sigma_2^2)}{(\mu_1 - \mu_2)^2} \log\left(\frac{1}{2.4\delta}\right)$$

and

$$w_*(\boldsymbol{\mu}) = \left[\frac{\sigma_1}{\sigma_1 + \sigma_2}; \frac{\sigma_2}{\sigma_1 + \sigma_2}\right]$$

➜ allocate the arms proportionaly to the standard deviations (no uniform sampling if $\sigma_1 \neq \sigma_2$)

**Optimal algorithm:**

- Sampling rule:

$$A_{t+1} = 1 \quad \Leftrightarrow \quad \frac{N_1(t)}{t} < \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

- Stopping rule: stop after $t$ samples if

$$|\hat{\mu}_1(t) - \hat{\mu}_2(t)| > \sqrt{2\left(\frac{\sigma_1^2}{N_1(t)} + \frac{\sigma_2^2}{N_2(t)}\right)\beta(t, \delta)}$$

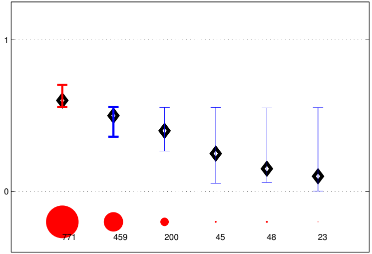
# Outline


1. Optimal algorithms for best-arm identification
   - Lower bounds
   - The Track-and-Stop strategy
2. A/B Testing
   - Bernoulli distribution
   - Gaussian distribution
3. Practical performance

# State-of-the-art algorithms

An algorithm based on confidence intervals : **KL-LUCB** [K., Kalyanakrishnan 13]

$$
\begin{aligned}
u_a(t) &= \max\left\{q : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\right\} \\
l_a(t) &= \min\left\{q : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\right\}
\end{aligned}
$$




- sampling rule: $A_{t+1} = \underset{a}{\operatorname{argmax}}\ \hat{\mu}_a(t)$, $B_{t+1} = \underset{b \neq A_{t+1}}{\operatorname{argmax}}\ u_b(t)$
- stopping rule: $\tau = \inf\{t \in \mathbb{N} : l_{A_t}(t) > u_{B_t}(t)\}$

# State-of-the-art algorithms

A Racing-type algorithm: **KL-Racing** [K., Kalyanakrishnan 13]

$\mathcal{R} = \{1, \ldots, K\}$ set of remaining arms.
$r = 0$ current round

**while** $|\mathcal{R}| > 1$

- r=r+1
- draw each $a \in \mathcal{R}$, compute $\hat{\mu}_{a,r}$, the empirical mean of the r samples observed sofar
- compute the empirical best and empirical worst arms:

$$b_r = \underset{a \in \mathcal{R}}{\text{argmax}}\ \hat{\mu}_{a,r} \qquad w_r = \underset{a \in \mathcal{R}}{\text{argmin}}\ \hat{\mu}_{a,r}$$

- Elimination step: if

$$l_{b_r}(r) > u_{w_r}(r),$$

eliminate $w_r$ : $\mathcal{R} = \mathcal{R} \backslash \{w_r\}$

**end**

**Outpout**: $\hat{a}$ the single element in $\mathcal{R}$.

# The Chernoff-Racing algorithm

$\mathcal{R} = \{1, \dots, K\}$ set of remaining arms.

$r = 0$ current round

**while** $|\mathcal{R}| > 1$

- r=r+1
- draw each $a \in \mathcal{R}$, compute $\hat{\mu}_{a,r}$, the empirical mean of the r samples observed sofar
- compute the empirical best and empirical worst arms:

$$b_r = \underset{a \in \mathcal{R}}{\text{argmax}} \ \hat{\mu}_{a,r} \qquad w_r = \underset{a \in \mathcal{R}}{\text{argmin}} \ \hat{\mu}_{a,r}$$

- Elimination step: if $(Z_{b_r,w_r}(r) > \beta(r,\delta))$, or

$$rd\left(\hat{\mu}_{a,r}, \frac{\hat{\mu}_{a,r} + \hat{\mu}_{b,r}}{2}\right) + rd\left(\hat{\mu}_{b,r}, \frac{\hat{\mu}_{a,r} + \hat{\mu}_{b,r}}{2}\right) > \beta(r,\delta),$$

  eliminate $w_r$ : $\mathcal{R} = \mathcal{R} \backslash \{w_r\}$

**end**

**Outpout**: $\hat{a}$ the single element in $\mathcal{R}$.

# Numerical experiments

Experiments on two Bernoulli bandit models:

- $\boldsymbol{\mu}_1 = [0.5\ 0.45\ 0.43\ 0.4]$, such that

$$w^*(\boldsymbol{\mu}_1) = [0.417\ 0.390\ 0.136\ 0.057]$$

- $\boldsymbol{\mu}_2 = [0.3\ 0.21\ 0.2\ 0.19\ 0.18]$, such that

$$w^*(\boldsymbol{\mu}_2) = [0.336\ 0.251\ 0.177\ 0.132\ 0.104]$$

In practice, set the threshold to $\beta(t,\delta) = \log\left(\frac{\log(t)+1}{\delta}\right)$.

| | Track-and-Stop | Chernoff-Racing | KL-LUCB | KL-Racing |
|---|---|---|---|---|
| $\boldsymbol{\mu}_1$ | 4052 | 4516 | 8437 | 9590 |
| $\boldsymbol{\mu}_2$ | 1406 | 3078 | 2716 | 3334 |

Table : Expected number of draws $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ for $\delta = 0.1$, averaged over $N = 3000$ experiments.

# Take-home message

Useful tools for sequential A/B Testing:

- stop using Sequential Generalized Likelihood Ratio tests
- sample the arms to match the optimal proportions $w^*(\boldsymbol{\mu})$
- ... which can be approximated by uniform sampling for Bernoulli distribution

**Final remark:**

Good algorithms for best arm identification are very different for bandit algorithms designed for regret minimization (UCB, Thompson Sampling)

# References

This talk is based on


- A. Garivier, E. Kaufmann.
  *Optimal Best Arm Identification with Fixed Confidence*,
  arXiv:1602.04589, 2016
- E. Kaufmann, O. Cappé, A. Garivier.
  *On the Complexity of A/B Testing*. COLT, 2014
- E. Kaufmann, O. Cappé, A. Garivier.
  *On the Complexity of Best Arm Identification in Multi-Armed Bandit Models*. JMLR, 2015