

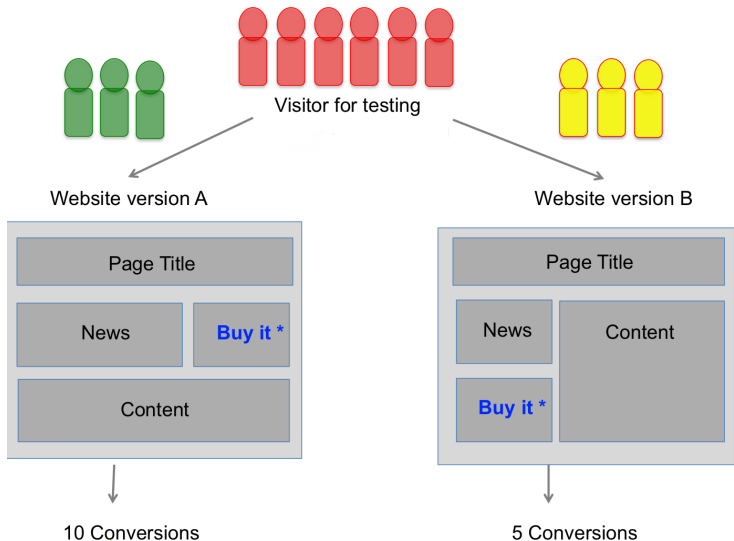
On the Complexity of A/B Testing

Emilie Kaufmann, Olivier Cappé (Télécom ParisTech) and Aurélien Garivier (Institut de Mathématiques de Toulouse)



Conference On Learning Theory, Barcelona,
June 14th, 2014

Motivation



Our goal

Improve performance:

- fixed number of test users – > smaller probability of error
- fixed probability of error – > fewer test users

Tools: sequential allocation and stopping

- 1 Best arm identification in two-armed bandits
- 2 Lower bounds on the complexities
- 3 The complexity of A/B Testing with Gaussian feedback
- 4 The complexity of A/B Testing with binary feedback

The model

A two-armed bandit model is

- a set $\nu = (\nu_1, \nu_2)$ of two probability distributions ('arms') with respective means μ_1 and μ_2
- $a^* = \operatorname{argmax}_a \mu_a$ is the (unknown) best arm

To find the best arm, an agent interacts with the bandit model with

- a *sampling rule* $(A_t)_{t \in \mathbb{N}}$ where $A_t \in \{1, 2\}$ is the arm chosen at time t (based on past observations) \rightarrow a sample $Z_t \sim \nu_{A_t}$ is observed
- a *stopping rule* τ indicating when he stops sampling the arms
- a *recommendation rule* $\hat{a}_\tau \in \{1, 2\}$ indicating which arm he thinks is best (at the end of the interaction)

In classical A/B Testing, the sampling rule A_t is uniform on $\{1, 2\}$ and the stopping rule $\tau = t$ is fixed in advance.

Two possible goals

The agent's goal is to design a strategy $\mathcal{A} = ((A_t), \tau, \hat{a}_\tau)$ satisfying

Fixed-budget setting	Fixed-confidence setting
$\tau = t$	$\mathbb{P}_\nu(\hat{a}_\tau \neq a^*) \leq \delta$
$p_t(\nu) := \mathbb{P}_\nu(\hat{a}_t \neq a^*)$ as small as possible	$\mathbb{E}_\nu[\tau]$ as small as possible

An algorithm using **uniform sampling** is

Fixed-budget setting	Fixed-confidence setting
a classical test of $(\mu_1 > \mu_2)$ against $(\mu_1 < \mu_2)$ based on t samples	a sequential test of $(\mu_1 > \mu_2)$ against $(\mu_1 < \mu_2)$ with probability of error uniformly bounded by δ

[Siegmund 85]: sequential tests can save samples !

The complexities of best-arm identification

Let \mathcal{M} be a class of bandit models. An algorithm $\mathcal{A} = ((A_t), \tau, \hat{a}_\tau)$ is...

Fixed-budget setting	Fixed-confidence setting
<p>consistent on \mathcal{M} if</p> $\forall \nu \in \mathcal{M}, p_t(\nu) = \mathbb{P}_\nu(\hat{a}_t \neq a^*) \xrightarrow[t \rightarrow \infty]{} 0$	<p>δ-PAC on \mathcal{M} if</p> $\forall \nu \in \mathcal{M}, \mathbb{P}_\nu(\hat{a}_\tau \neq a^*) \leq \delta$

From the literature

$p_t(\nu) \simeq \exp\left(-\frac{t}{CH(\nu)}\right)$ <p>[Audibert et al. 10],[Bubeck et al. 11] [Bubeck et al. 13],...</p>	$\mathbb{E}_\nu[\tau] \simeq C' H'(\nu) \log \frac{1}{\delta}$ <p>[Mannor Tsitsilis 04],[Even-Dar et al. 06] [Kalanakrishnan et al.12],...</p>
---	--

Two complexities

$\kappa_B(\nu) = \inf_{\mathcal{A} \text{ consistent}} \left(\limsup_{t \rightarrow \infty} -\frac{1}{t} \log p_t(\nu) \right)^{-1}$ <p>for a probability of error $\leq \delta$, budget $t \simeq \kappa_B(\nu) \log \frac{1}{\delta}$</p>	$\kappa_C(\nu) = \inf_{\mathcal{A} \delta\text{-PAC}} \limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)}$ <p>for a probability of error $\leq \delta$ $\mathbb{E}_\nu[\tau] \simeq \kappa_C(\nu) \log \frac{1}{\delta}$</p>
--	---

Outline

- 1 Best arm identification in two-armed bandits
- 2 Lower bounds on the complexities**
- 3 The complexity of A/B Testing with Gaussian feedback
- 4 The complexity of A/B Testing with binary feedback

Changes of distribution

New formulation for a change of distribution

Let ν and ν' be two bandit models. Let N_1 (resp. N_2) denote the total number of draws of arm 1 (resp. arm 2) by algorithm \mathcal{A}). For any $A \in \mathcal{F}_\tau$ such that $0 < \mathbb{P}_\nu(A) < 1$

$$\mathbb{E}_\nu[N_1] \text{KL}(\nu_1, \nu'_1) + \mathbb{E}_\nu[N_2] \text{KL}(\nu_2, \nu'_2) \geq d(\mathbb{P}_\nu(A), \mathbb{P}_{\nu'}(A)),$$

where $d(x, y) := x \log(x/y) + (1 - x) \log((1 - x)/(1 - y))$.

General lower bounds

Theorem 1

Let \mathcal{M} be a class of two armed bandit models that are continuously parametrized by their means. Let $\nu = (\nu_1, \nu_2) \in \mathcal{M}$.

Fixed-budget setting	Fixed-confidence setting
any consistent algorithm satisfies	any δ -PAC algorithm satisfies
$\limsup_{t \rightarrow \infty} -\frac{1}{t} \log p_t(\nu) \leq K^*(\nu_1, \nu_2)$ <p style="text-align: center;">with $KL^*(\nu_1, \nu_2)$ $= KL(\nu^*, \nu_1) = KL(\nu^*, \nu_2)$</p>	$\mathbb{E}_\nu[\tau] \geq \frac{1}{K_*(\nu_1, \nu_2)} \log\left(\frac{1}{2\delta}\right)$ <p style="text-align: center;">with $K_*(\nu_1, \nu_2)$ $= KL(\nu_1, \nu_*) = K(\nu_2, \nu_*)$</p>
Thus, $\kappa_B(\nu) \geq \frac{1}{K^*(\nu_1, \nu_2)}$	Thus, $\kappa_C(\nu) \geq \frac{1}{K_*(\nu_1, \nu_2)}$

- 1 Best arm identification in two-armed bandits
- 2 Lower bounds on the complexities
- 3 The complexity of A/B Testing with Gaussian feedback
- 4 The complexity of A/B Testing with binary feedback

Fixed-budget setting

For fixed (known) values σ_1, σ_2 , we consider Gaussian bandit models

$$\mathcal{M} = \left\{ \nu = \left(\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2) \right) : (\mu_1, \mu_2) \in \mathbb{R}^2, \mu_1 \neq \mu_2 \right\}$$

■ Theorem 1:

$$\kappa_B(\nu) \geq \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}$$

- A strategy allocating $t_1 = \left\lfloor \frac{\sigma_1}{\sigma_1 + \sigma_2} t \right\rfloor$ samples to arm 1 and $t_2 = t - t_1$ samples to arm 2, and recommending the empirical best satisfies

$$\liminf_{t \rightarrow \infty} -\frac{1}{t} \log p_t(\nu) \geq \frac{(\mu_1 - \mu_2)^2}{2(\sigma_1 + \sigma_2)^2}$$

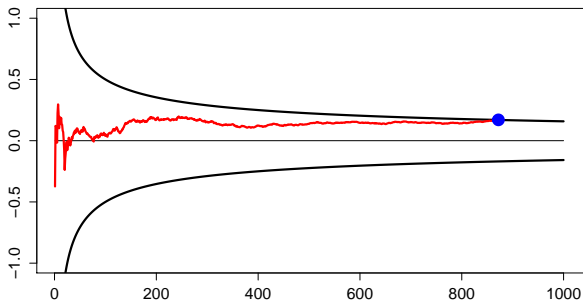
$$\kappa_B(\nu) = \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}$$

Fixed-confidence setting: Algorithm

The α -Elimination algorithm with exploration rate $\beta(t, \delta)$

- chooses A_t in order to keep a proportion $N_1(t)/t \simeq \alpha$
i.e. $A_t = 2$ if and only if $\lceil \alpha t \rceil = \lceil \alpha(t+1) \rceil$
- if $\hat{\mu}_a(t)$ is the empirical mean of rewards obtained from a up to time t , $\sigma_t^2(\alpha) = \sigma_1^2/\lceil \alpha t \rceil + \sigma_2^2/(t - \lceil \alpha t \rceil)$,

$$\tau = \inf \left\{ t \in \mathbb{N} : |\hat{\mu}_1(t) - \hat{\mu}_2(t)| > \sqrt{2\sigma_t^2(\alpha)\beta(t, \delta)} \right\}$$



Fixed-confidence setting: Results

- From Theorem 1:

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

- $\frac{\sigma_1}{\sigma_1 + \sigma_2}$ -Elimination with $\beta(t, \delta) = \log \frac{t}{\delta} + 2 \log \log(6t)$ is δ -PAC and

$$\forall \epsilon > 0, \quad \mathbb{E}_\nu[\tau] \leq (1 + \epsilon) \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2} \log\left(\frac{1}{2\delta}\right) + o_{\delta \rightarrow 0} \left(\log \frac{1}{\delta}\right)$$

$$\kappa_C(\nu) = \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}$$

Gaussian distributions: Conclusions

For any two fixed values of σ_1 and σ_2 ,

$$\kappa_B(\nu) = \kappa_C(\nu) = \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}$$

If the variances are equal, $\sigma_1 = \sigma_2 = \sigma$,

$$\kappa_B(\nu) = \kappa_C(\nu) = \frac{8\sigma^2}{(\mu_1 - \mu_2)^2}$$

- **uniform sampling** is optimal only when $\sigma_1 = \sigma_2$
- 1/2-Elimination is δ -PAC for a smaller exploration rate
 $\beta(t, \delta) \simeq \log(\log(t)/\delta)$

- 1 Best arm identification in two-armed bandits
- 2 Lower bounds on the complexities
- 3 The complexity of A/B Testing with Gaussian feedback
- 4 The complexity of A/B Testing with binary feedback

Lower bounds for Bernoulli bandit models

$$\mathcal{M} = \{\nu = (\mathcal{B}(\mu_1), \mathcal{B}(\mu_2)) : (\mu_1, \mu_2) \in]0; 1[^2, \mu_1 \neq \mu_2\},$$

shorthand: $K(\mu, \mu') = \text{KL}(\mathcal{B}(\mu), \mathcal{B}(\mu'))$.

Fixed-budget setting	Fixed-confidence setting
any consistent algorithm satisfies	any δ -PAC algorithm satisfies
$\limsup_{t \rightarrow \infty} -\frac{1}{t} \log p_t(\nu) \leq K^*(\mu_1, \mu_2)$	$\mathbb{E}_\nu[\tau] \geq \frac{1}{K_*(\mu_1, \mu_2)} \log\left(\frac{1}{2\delta}\right)$
(Chernoff information)	

$$K^*(\mu_1, \mu_2) > K_*(\mu_1, \mu_2)$$

Algorithms using uniform sampling

	For any consistent...	For any δ -PAC...
... algorithm	$p_t(\nu) \gtrsim e^{-K^*(\mu_1, \mu_2)t}$	$\frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \gtrsim \frac{1}{K_*(\mu_1, \mu_2)}$
... algorithm using uniform sampling	$p_t(\nu) \gtrsim e^{-\frac{K(\bar{\mu}, \mu_1) + K(\bar{\mu}, \mu_2)}{2}t}$ with $\bar{\mu} = f(\mu_1, \mu_2)$	$\frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \gtrsim \frac{2}{K(\mu_1, \underline{\mu}) + K(\mu_2, \underline{\mu})}$ with $\underline{\mu} = \frac{\mu_1 + \mu_2}{2}$

Remark: Quantities in the same column appear to be close from one another

⇒ **Binary rewards: uniform sampling close to optimal**

Algorithms using uniform sampling

	For any consistent...	For any δ -PAC...
... algorithm	$p_t(\nu) \simeq e^{-K^*(\mu_1, \mu_2)t}$	$\frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \gtrsim \frac{1}{K_*(\mu_1, \mu_2)}$
... algorithm using uniform sampling	$p_t(\nu) \simeq e^{-\frac{K(\bar{\mu}, \mu_1) + K(\bar{\mu}, \mu_2)}{2}t}$ with $\bar{\mu} = f(\mu_1, \mu_2)$	$\frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \gtrsim \frac{2}{K(\mu_1, \underline{\mu}) + K(\mu_2, \underline{\mu})}$ with $\underline{\mu} = \frac{\mu_1 + \mu_2}{2}$

Remark: Quantities in the same column appear to be close from one another

⇒ **Binary rewards: uniform sampling close to optimal**

Fixed-budget setting

We show that

$$\kappa_B(\nu) = \frac{1}{\mathbf{K}^*(\mu_1, \mu_2)}$$

(matching algorithm not implementable in practice)

The algorithm using **uniform sampling** and recommending the empirical best arm **is preferable** (and very close to optimal)

Fixed-confidence setting

δ -PAC algorithms using uniform sampling satisfy

$$\frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \geq \frac{1}{I_*(\nu)} \quad \text{with} \quad I_*(\nu) = \frac{K\left(\mu_1, \frac{\mu_1 + \mu_2}{2}\right) + K\left(\mu_2, \frac{\mu_1 + \mu_2}{2}\right)}{2}.$$

The algorithm using uniform sampling and

$$\tau = \inf \left\{ t \in 2\mathbb{N}^* : |\hat{\mu}_1(t) - \hat{\mu}_2(t)| > \log \frac{\log(t) + 1}{\delta} \right\}$$

is δ -PAC but not optimal: $\frac{\mathbb{E}[\tau]}{\log(1/\delta)} \simeq \frac{2}{(\mu_1 - \mu_2)^2} > \frac{1}{I_*(\nu)}$.

A better stopping rule NOT based on the difference of empirical means

$$\tau = \inf \left\{ t \in 2\mathbb{N}^* : t I_*(\hat{\mu}_1(t), \hat{\mu}_2(t)) > \log \frac{\log(t) + 1}{\delta} \right\}$$

Bernoulli distributions: Conclusion

Regarding the complexities:

- $\kappa_B(\nu) = \frac{1}{K^*(\mu_1, \mu_2)}$
- $\kappa_C(\nu) \geq \frac{1}{K_*(\mu_1, \mu_2)} > \frac{1}{K^*(\mu_1, \mu_2)}$

Thus

$$\kappa_C(\nu) > \kappa_B(\nu)$$

Regarding the algorithms

- There is not much to gain by departing from uniform sampling
- In the fixed-confidence setting, a sequential test based on the difference of the empirical means is no longer optimal

Conclusion

- the complexities $\kappa_B(\nu)$ and $\kappa_C(\nu)$ are not always equal (and feature some different informational quantities)
- for Bernoulli distributions and Gaussian with similar variances, strategies using uniform sampling are (almost) optimal
- strategies using random stopping do not necessarily lead to a saving in terms of the number of sample used

Coming soon:

- Generalization to m best arms identification among K arms