I'm a bandit
Random topics in optimization, probability, and statistics. By Sébastien Bubeck

Bandit theory, part I

Posted on May 11, 2016 by Sebastien Bubeck

This week I'm giving two 90 minutes lectures on bandit theory at MLSS Cadiz. Despite my 2012 survey with Nicolo I thought it would be a good idea to post my lectures notes here. Indeed while much of the material is similar, the style of a mini-course is quite different from the style of a survey. Also, bandit theory has surprisingly progressed since 2012 and many things can now be explained better. Finally in the survey we completely omitted the Bayesian model as we thought that we didn't have much to add on this topic compared to existing sources (such as the 2011 book by Gittins, Glazebrook, Weber). For a mini-course this concern is irrelevant so I quickly discuss the famous Gittins index and its proof of optimality.

i.i.d. multi-armed bandit, Robbins [1952]

Known parameters: number of arms n and (possibly) number of rounds $T \geq n$.

Unknown parameters: n probability distributions $\nu_1, \dots, \nu_n$ on $[0,1]$ with mean $\mu_1, \dots, \mu_n$ (notation: $\mu^* = \max_{i \in [n]} \mu_i$).

Protocol: For each round $t = 1, 2, \dots, T$, the player chooses $I_t \in [n]$ based on past observations and receives a reward/observation $Y_t \sim \nu_{I_t}$ (independently from the past).

Performance measure: The cumulative regret is the difference between the player's accumulated reward and the maximum the player could have obtained had she known all the parameters,

$$\overline{R}_T = T\mu^* - \mathbb{E} \sum_{t \in [T]} Y_t .$$

This problem models the fundamental tension between **exploration** and **exploitation** (one wants to pick arms that performed well in the past, yet one needs to make sure that no good option has been missed). Almost every week new applications are found that fit this simple framework and I'm sure you already have some in mind (the most popular one being ad placement on the internet).

i.i.d. multi-armed bandit: fundamental limitations

How small can we expect $\overline{R}_T$ to be? Consider the 2-armed case where $\nu_1 = Ber(1/2)$ and $\nu_2 = Ber(1/2 + \xi\Delta)$ where $\xi \in \{-1, 1\}$ is unknown. Recall from Probability 101 (or perhaps 102) that with τ expected observations from the second arm there is a probability at least $\exp(-\tau\Delta^2)$ to make the wrong guess on the value of ξ. Now let $\tau(t)$ be the expected number of pulls of arm 2 up to time t when $\xi = -1$. One has

$$\begin{aligned}\overline{R}_T(\xi = +1) + \overline{R}_T(\xi = -1) & \geq \Delta\tau(T) + \Delta \sum_{t=1}^{T} \exp(-\tau(t)\Delta^2) \\ & \geq \Delta \min_{t \in [T]} (t + T\exp(-t\Delta^2)) \\ & \approx \frac{\log(T\Delta^2)}{\Delta}.\end{aligned}$$

We refer to Bubeck, Perchet and Rigollet [2013] for the details. The important message is that for Δ fixed the lower bound is $\frac{\log(T)}{\Delta}$, while for the worse Δ (which is of order $1/\sqrt{T}$) it is $\sqrt{T}$. In the n-armed case this worst-case lower bound becomes $\sqrt{Tn}$ (see Auer, Cesa-Bianchi, Freund and Schapire [1995]). The $\log(T)$-lower bound is slightly "harder" to generalize to the n-armed case (as far as I know there is no known finite-time lower bound of this type), but thankfully it was already all done 30 years ago. First some notation: let $\Delta_i = \mu^* - \mu_i$ and $N_i(t)$ the number of pulls of arm i up to time t. Note that one has $\overline{R}_T = \sum_{i=1}^n \Delta_i \mathbb{E} N_i(T)$. For $p, q \in [0, 1]$ let

$$\mathrm{kl}(p, q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$

> **Theorem** [Lai and Robbins [1985]]
>
> Consider a strategy s.t. $\forall a > 0$, we have $\mathbb{E} N_i(T) = o(T^a)$ if $\Delta_i > 0$. Then for any Bernoulli distributions,
>
> $$\liminf_{T \to +\infty} \frac{\overline{R}_T}{\log(T)} \geq \sum_{i : \Delta_i > 0} \frac{\Delta_i}{\mathrm{kl}(\mu_i, \mu^*)}.$$

Note that $\frac{1}{2\Delta_i} \geq \frac{\Delta_i}{\mathrm{kl}(\mu_i, \mu^*)} \geq \frac{\mu^*(1-\mu^*)}{2\Delta_i}$ so up to a variance-like term the Lai and Robbins lower bound is $\sum_{i:\Delta_i>0} \frac{\log(T)}{2\Delta_i}$. This lower bound holds more generally than just for Bernoulli distributions, see for example Burnetas and Katehakis [1996].

i.i.d. multi-armed bandit: fundamental strategy

Hoeffding's inequality teaches us that with probability at least $1 - 1/T$, $\forall t \in [T], i \in [n]$,

$$\mu_i \leq \frac{1}{N_i(t)} \sum_{s<t : I_s = i} Y_s + \sqrt{\frac{2\log(T)}{N_i(t)}} =: \mathrm{UCB}_i(t).$$

The UCB (Upper Confidence Bound) strategy (Lai and Robbins [1985], Agarwal [1995], Auer, Cesa-Bianchi and Fischer [2002]) is:

$$I_t \in \mathrm{argmax}_{i \in [n]} \mathrm{UCB}_i(t).$$

The regret analysis is straightforward: on a $1 - 2/T$ probability event one has

$$N_i(t) \geq 8\log(T)/\Delta_i^2 \Rightarrow \mathrm{UCB}_i(t) < \mu^* \leq \mathrm{UCB}_{i^*}(t),$$

so that $\mathbb{E} N_i(T) \leq 2 + 8 \log(T)/\Delta_i^2$ and in fact

$$\overline{R}_T \leq 2 + \sum_{i : \Delta_i > 0} \frac{8 \log(T)}{\Delta_i}.$$

i.i.d. multi-armed bandit: going further

- The numerical constant 8 in the UCB regret bound can be replaced by $1/2$ (which is the best one can hope for), and more importantly by slightly modifying the derivation of the UCB one can obtain the Lai and Robbins variance-like term (that is replacing Δ_i by $\mathrm{kl}(\mu_i, \mu^*)$): see Cappe, Garivier, Maillard, Munos and Stoltz [2013].
- In many applications one is merely interested in *finding* the best arm (instead of maximizing cumulative reward): this is the best arm identification problem. For the fundamental strategies see Even-Dar, Mannor and Mansour [2006] for the fixed-confidence setting (see also Jamieson and Nowak [2014] for a recent short survey) and Audibert, Bubeck and Munos [2010] for the fixed budget setting. Key takeaway: one needs of order $\mathbf{H} := \sum_i \Delta_i^{-2}$ rounds to find the best arm.
- The UCB analysis extends to sub-Gaussian reward distributions. For heavy-tailed distributions, say with $1+\epsilon$ moment for some $\epsilon \in (0,1]$, one can get a regret that scales with $\Delta_i^{-1/\epsilon}$ (instead of Δ_i^{-1}) by using a robust mean estimator, see Bubeck, Cesa-Bianchi and Lugosi [2012].

Adversarial multi-armed bandit, Auer, Cesa-Bianchi, Freund and Schapire [1995, 2001]

For $t = 1, \dots, T$, the player chooses $I_t \in [n]$ based on previous observations, and simultaneously an adversary chooses a loss vector $\ell_t \in [0,1]^n$. The player's loss/observation is $\ell_t(I_t)$. The regret and pseudo-regret are defined as:

$$R_T = \max_{i \in [n]} \sum_{t \in [T]} (\ell_t(I_t) - \ell_t(i)), \quad \overline{R}_T = \max_{i \in [n]} \mathbb{E} \sum_{t \in [T]} (\ell_t(I_t) - \ell_t(i)).$$

Obviously $\mathbb{E} R_T \geq \overline{R}_T$ and there is equality in the oblivious case ($\equiv$ adversary's choices are independent of the player's choices). The case where $\ell_1, \dots, \ell_T$ is an i.i.d. sequence corresponds to the i.i.d. model we just studied. In particular we already know that $\sqrt{Tn}$ is a lower bound on the attainable pseudo-regret.

Adversarial multi-armed bandit, fundamental strategy

The exponential weights strategy for the *full information* case where ℓ_t is observed at the end of round t is defined by: play I_t at random from p_t where

$$p_{t+1}(i) = \frac{1}{Z_{t+1}} p_t(i) \exp(-\eta \ell_t(i)).$$

In five lines one can show $\overline{R}_T \leq \sqrt{2T \log(n)}$ with $p_1(i) = 1/n$ and a well-chosen learning rate η (recall that $\mathrm{Ent}(p \| q) = \sum_i p(i) \log(p(i)/q(i))$):

$$\mathrm{Ent}(\delta_j \| p_t) - \mathrm{Ent}(\delta_j \| p_{t+1}) = \log \frac{p_{t+1}(j)}{p_t(j)} = \log \frac{1}{Z_{t+1}} - \eta \ell_t(j)$$

$$\psi_t := \log \mathbb{E}_{I \sim p_t} \exp(-\eta(\ell_t(I) - \mathbb{E}_{I' \sim p_t} \ell_t(I'))) = \eta \mathbb{E} \ell_t(I') + \log(Z_{t+1})$$

$$\eta \sum_t \left(\sum_i p_t(i) \ell_t(i) - \ell_t(j) \right) = \mathrm{Ent}(\delta_j \| p_1) - \mathrm{Ent}(\delta_j \| p_{T+1}) + \sum_t \psi_t$$

$$\text{Using that } \ell_t \geq 0 \text{ one has } \psi_t \leq \frac{\eta^2}{2} \mathbb{E} \ell_t(i)^2 \text{ thus } \overline{R}_T \leq \frac{\log(n)}{\eta} + \frac{\eta T}{2}$$

For the bandit case we replace ℓ_t by $\tilde{\ell}_t$ in the exponential weights strategy, where

$$\tilde{\ell}_t(i) = \frac{\ell_t(I_t)}{p_t(i)} 1\{i = I_t\}.$$

The resulting strategy is called Exp3. The key property of $\tilde{\ell}_t$ is that it is an unbiased estimator of ℓ_t:

$$\mathbb{E}_{I_t \sim p_t} \tilde{\ell}_t(i) = \ell_t(i).$$

Furthermore with the analysis described above one gets

$$\overline{R}_T \leq \frac{\log(n)}{\eta} + \frac{\eta}{2} \mathbb{E} \sum_t \mathbb{E}_{I \sim p_t} \tilde{\ell}_t(I)^2.$$

It only remains to control the variance term, and quite amazingly this is straightforward:

$$\mathbb{E}_{I_t, I \sim p_t} \tilde{\ell}_t(I)^2 \leq \mathbb{E}_{I_t, I \sim p_t} \frac{1\{I = I_t\}}{p_t(I_t)^2} = n.$$

Thus with $\eta = \sqrt{2n \log(n)/T}$ one gets $\overline{R}_T \leq \sqrt{2Tn \log(n)}$.

Adversarial multi-armed bandit, going further

- With the modified loss estimate $\frac{\ell_t(I_t)1\{i=I_t\}+\beta}{p_t(I_t)}$ one can prove high probability bounds on R_T, and by integrating the deviations one can show $\mathbb{E} R_T = O(\sqrt{Tn \log(n)})$.
- The extraneous logarithmic factor in the pseudo-regret upper can be removed, see Audibert and Bubeck [2009]. Conjecture: one cannot remove the log factor for the expected regret, that is for any strategy there exists an adaptive adversary such that $\mathbb{E} R_T = \Omega(\sqrt{Tn \log(n)})$.
- T can be replaced by various measure of "variance" in the loss sequence, see e.g., Hazan and Kale [2009].
- There exist strategies which guarantee simultaneously $\overline{R}_T = \tilde{O}(\sqrt{Tn})$ in the adversarial model and $\overline{R}_T = \tilde{O}(\sum_i \Delta_i^{-1})$ in the i.i.d. model, see Bubeck and Slivkins [2012].
- Many interesting variants: graph feedback structure of Mannor and Shamir [2011] (there is a graph on the set of arms, and when an arm is played one observes the loss for all its neighbors), regret with respect to best sequence of actions with at most S switches, switching cost (interestingly in this case the best regret is $\Omega(T^{2/3})$, see Dekel, Ding, Koren and Peres [2013]), and much more!

Bayesian multi-armed bandit, Thompson [1933]

Here we assume a set of "models" $\{(\nu_1(\theta), \dots, \nu_n(\theta)), \theta \in \Theta\}$ and prior distribution π_0 over Θ. The Bayesian regret is defined as

$$BR_T(\pi_0) = \mathbb{E}_{\theta \sim \pi_0} \overline{R}_T(\nu_1(\theta), \dots, \nu_n(\theta)),$$

where $\overline{R}_T(\nu)$ simply denotes the regret for the i.i.d. model when the underlying reward distributions are $\nu_1, \dots, \nu_n$. In principle the strategy minimizing the Bayesian regret can be computed by dynamic programming on the potentially huge state space $\mathcal{P}(\Theta)$. The celebrated Gittins index theorem gives sufficient condition to dramatically reduce the computational complexity of implementing the optimal Bayesian strategy under a strong product assumption on π_0. Notation: π_t denotes the posterior distribution on θ at time t.

> **Theorem** [Gittins [1979]]
>
> Consider the product and γ-discounted case: $\Theta = \times_i \Theta_i$, $\nu_i(\theta) := \nu(\theta_i)$, $\pi_0 = \otimes_i \pi_0(i)$, and furthermore one is interested in maximizing $\mathbb{E} \sum_{t>0} \gamma^t Y_t$. The optimal Bayesian strategy is to pick at time s the arm maximizing the *Gittins index*:

$$\sup\left\{\lambda \in \mathbb{R} : \sup_{\tau} \mathbb{E}\left(\sum_{t<\tau} \gamma^t X_t + \frac{\gamma^\tau}{1-\gamma}\lambda\right) \geq \frac{1}{1-\gamma}\lambda\right\},$$

> where the expectation is over (X_t) drawn from $\nu(\theta)$ with $\theta \sim \pi_s(i)$, and the supremum is taken over all stopping times τ.

Note that the stopping time τ in the Gittins index definition gives the optimal strategy for a 2-armed game, where one arm's reward distribution is δ_λ while the other arm reward's distribution is $\nu(\theta)$ with $\pi_s(i)$ as a prior for θ.

Proof: The following exquisite proof was discovered by Weber [1992]. Let

$$\lambda_t(i) := \sup\left\{\lambda \in \mathbb{R} : \sup_{\tau} \mathbb{E} \sum_{t<\tau} \gamma^t (X_t - \lambda) \geq 0\right\}$$

be the Gittins index of arm i at time t, which we interpret as the *maximum charge* one is willing to pay to play arm i given the current information. The *prevailing charge* is defined as $\overline{\lambda}_t(i) = \min_{s \leq t} \lambda_s(i)$ (i.e. whenever the prevailing charge is too high we just drop it to the fair level). We now make three simple observations which together conclude the proof:

- The discounted sum of prevailing charge for played arms $\sum_t \gamma^t \overline{\lambda}_t(I_t)$ is an upper bound (in expectation) on the discounted sum of rewards. Indeed the times at which the prevailing charge are updated are stopping times, and so between two such times (s, t) the expected sum of discounted reward is smaller than the discounted sum of the fair charge at time s which is equal to the prevailing charge at any time in $[s, t-1]$.
- Since the prevailing charge is nonincreasing, the discounted sum of prevailing charge is maximized if we always pick the arm with maximum prevailing charge. Also note that the sequence of prevailing charge $(\overline{\lambda}_t(i))_{i,t}$ does not depend on the algorithm.
- Gittins index does exactly 2. (since we stop playing an arm only at times at which the prevailing charge is updated) and in this case 1. is an equality. Q.E.D.

For much more (implementation for exponential families, interpretation as a multitoken Markov game, …) see Dumitriu, Tetali and Winkler [2003], Gittins, Glazebrook, Weber [2011], Kaufmann [2014].

Bayesian multi-armed bandit, Thompson Sampling (TS)

In machine learning we want (i) strategies that can deal with complicated priors, and (ii) guarantees for misspecified priors. This is why we have to go beyond the Gittins index theory.

In his 1933 paper Thompson proposed the following strategy: sample $\theta' \sim \pi_t$ and play $I_t \in \mathrm{argmax} \mu_i(\theta')$.

Theoretical guarantees for this highly practical strategy have long remained elusive. Recently Agrawal and Goyal [2012] and Kaufmann, Korda and Munos [2012] proved that TS with Bernoulli reward distributions and uniform prior on the parameters achieves $\overline{R}_T = O\left(\sum_i \frac{\log(T)}{\Delta_i}\right)$ (note that this is the frequentist regret!). We also note that Liu and Li [2015] takes some steps in analyzing the effect of misspecification for TS.

Let me also mention a beautiful conjecture of Guha and Munagala [2014]: for product priors, TS is a 2-approximation to the optimal Bayesian strategy for the objective of minimizing the number of pulls on suboptimal arms.

Bayesian multi-armed bandit, Russo and Van Roy [2014] information ratio analysis

Assume a prior in the adversarial model, that is a prior over $(\ell_1, \dots, \ell_T) \in [0,1]^{n \times T}$, and let $\mathbb{E}_t$ denote the posterior distribution (given $\ell_1(I_1), \dots, \ell_{t-1}(I_{t-1})$). We introduce

$$r_t(i) = \mathbb{E}_t(\ell_t(i) - \ell_t(i^*)), \ \text{ and } \ v_t(i) = \mathrm{Var}_t(\mathbb{E}_t(\ell_t(i) | i^*)).$$

The key observation is that (recall that $H(p) = -\sum_i p(i) \log(p(i))$)

$$\mathbb{E} \sum_{t \leq T} v_t(I_t) \leq \frac{1}{2} H(i^*)$$

Indeed, equipped with Pinsker’s inequality and basic information theory concepts one has (we denote I_t for the mutual information conditionally on everything up to time t, also $\mathcal{L}_t(X)$ denotes the law of X conditionally on everything up to time t):

$$\begin{aligned} v_t(i) & = \sum_j \pi_t(j) (\mathbb{E}_t(\ell_t(i) | i^* = j) - \mathbb{E}_t(\ell_t(i)))^2 \\ & \leq \frac{1}{2} \sum_j \pi_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) || \mathcal{L}_t(\ell_t(i))) \\ & = \frac{1}{2} I_t(\ell_t(i), i^*) = \frac{1}{2} (H_t(i^*) - H_t(i^* | \ell_t(i))). \end{aligned}$$

Thus $\mathbb{E} v_t(I_t) \leq \frac{1}{2} \mathbb{E} (H_t(i^*) - H_{t+1}(i^*))$ which gives the claim thanks to a telescopic sum. We will use this key observation as follows (we give a sequence of implications leading to a regret bound so that all that is left to do is to check that the first statement in this sequence is true for TS):

$$\begin{aligned} & \forall t, \mathbb{E}_t r_t(I_t) \leq \sqrt{C \ \mathbb{E}_t v_t(I_t)} \\ \Rightarrow & \ \ \mathbb{E} \sum_{t=1}^T r_t(I_t) \leq \sum_{t=1}^T \sqrt{C \ \mathbb{E} v_t(I_t)} \\ \Rightarrow & \ \ BR_T \leq \sqrt{C \ T \ H(i^*)/2}. \end{aligned}$$

Thus writing $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i) | i^* = j)$ we have

$$\begin{aligned} & \mathbb{E}_t r_t(I_t) \leq \sqrt{C \ \mathbb{E}_t v_t(I_t)} \\ \Leftrightarrow & \ \mathbb{E}_{I_t} \bar{\ell}_t(I_t) - \sum_i \pi_t(i) \bar{\ell}_t(i,i) \leq \sqrt{C \ \mathbb{E}_{I_t} \sum_j \pi_t(j) (\bar{\ell}_t(I_t, j) - \bar{\ell}_t(I_t))^2} \end{aligned}$$

For TS the following shows that one can take $C = n$:

$$\begin{aligned} \mathbb{E}_{I_t} \bar{\ell}_t(I_t) - \sum_i \pi_t(i) \bar{\ell}_t(i,i) & = \sum_i \pi_t(i) (\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \\ & \leq \sqrt{n \sum_i \pi_t(i)^2 (\bar{\ell}_t(i) - \bar{\ell}_t(i,i))^2} \\ & \leq \sqrt{n \sum_{i,j} \pi_t(i) \pi_t(j) (\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2}. \end{aligned}$$

Thus TS always satisfies $BR_T \leq \sqrt{T n H(i^*)} \leq \sqrt{T n \log(n)}$. Side note: by the minimax theorem this implies the existence of a strategy for the oblivious adversarial model with regret $\sqrt{T n \log(n)}$ (of course we already proved that such a strategy exist, in fact we even constructed one via exponential weights, but the point is that the proof here does not require any “miracle” –yes exponential weights are kind of a miracle, especially when you consider how the variance of the unbiased estimator gets automatically controlled).

Summary of basic results

- In the i.i.d. model UCB attains a regret of $O\left(\sum_i \frac{\log(T)}{\Delta_i}\right)$ and by Lai and Robbins' lower bound this is optimal (up to a multiplicative variance-like term).
- In the adversarial model Exp3 attains a regret of $O(\sqrt{Tn\log(n)})$ and this is optimal up to the logarithmic term.
- In the Bayesian model, Gittins index gives an *optimal* strategy for the case of product priors. For general priors Thompson Sampling is a more flexible strategy. Its Bayesian regret is controlled by the entropy of the optimal decision. Moreover TS with an uninformative prior has frequentist guarantees comparable to UCB.

This entry was posted in Optimization, Probability theory. Bookmark the permalink.
← COLT 2016 accepted papers
Bandit theory, part II →

6 Responses to "Bandit theory, part I"

- **By Jinshuo April 21, 2017 - 4:21 am**

  "there is equality in the oblivious case" is that really the case? When there are two arms and they have independent rewards(say X and Y), E max{X,Y} > max{EX, EY}

  Reply

- **By Adam Smith May 14, 2016 - 9:04 am**

  - "Now let \tau(t) be the expected number of pulls of arm 2 when \xi=-1." Did you mean pulls at steps 1,...,t? Right now t does not appear in the definition.

  - Missing "argmax" in the definition of the UCB strategy

  Reply

  - **By Sebastien Bubeck May 15, 2016 - 7:04 am**

    Fixed, thanks!

- **By Sebastien Bubeck May 12, 2016 - 10:52 pm**

Yes I agree David! This would have made a perfect exercise (if I had an exercise session ;). Another one would have been the 5-lines new proof of Lai and Robbins by Garivier, Menard and Stoltz http://arxiv.org/abs/1602.07182 . In any case for a 90 minutes what I have in the blog post was already slightly too much and I had to skip the Weber 1992 proof of optimality for Gittins index.

Reply

- **By David Pal May 11, 2016 - 5:22 pm**

  Abernethy et al. http://arxiv.org/abs/1512.04152 do an excellent job of explaining the optimal adversarial bounds.

  Reply

  - **By Sebastien Bubeck May 12, 2016 - 10:53 pm**

    Not sure why but I couldn't reply to your comment directly so instead I made another comment.

Leave a reply

Name

Email(will not be published)

Website

Submit comment

☐ Notify me of follow-up comments by email.

☐ Notify me of new posts by email.

- Search for: Search

Theme by

Simple Themes Princeton

- ## Archives

  Archives Select Month ▼

- ## Categories

  Categories Select Category ▼

- ## Recent Posts

  - k-server, part 3: entropy regularization for weighted k-paging
  - k-server, part 2: continuous time mirror descent
  - k-server, part 1: online learning and online algorithms
  - Algorithms, Machine Learning, and Optimization: we are hiring!
  - COLT 2018 call for papers

- ## Subscribe to Blog via Email

  Enter your email address to subscribe to this blog and receive notifications of new posts by email.

  Join 314 other subscribers

  Email Address

  Subscribe

- ## Meta

  - Log in

- RSS - Posts

  RSS - Comments

- ## Blogroll

  - Combinatorics and more
  - Computational Complexity
  - Godel's Lost Letter
  - Gowers's Weblog
  - hunch.net
  - in theory
  - Normal Deviate
  - Nuit Blanche
  - Shtetl-Optimized
  - Stochastic Analysis Seminar

- The Geomblog
- What's new

University
I’m a bandit

© 2018 The Trustees of Princeton University