Cosmic Bandits: Exploration versus Exploitation in Cosmological Surveys

Ely D. Kovetz University of Texas at Austin ITC Seminar, Dec. 10th, 2013





• How to balance the tradeoff between *exploring* and *exploiting*?

How to balance the tradeoff between *exploring* and *exploiting*?
One way...



How to balance the tradeoff between *exploring* and *exploiting*?
One way...



• Not necessarily stupid.

How to balance the tradeoff between *exploring* and *exploiting*?
One way...



- Not necessarily stupid.
- Different targets call for different measurements.

Stochastic measurements and deep-field imaging call for different approaches.

Stochastic measurements and deep-field imaging call for different approaches.

• Deep-field imaging:

Exploration mostly wasted.



(Hubble UDF)



Stochastic measurements and deep-field imaging call for different approaches.

- Deep-field imaging:
- Exploration mostly wasted.
- Goal of adaptive strategy:
 - --> quickly converge and exploit.



(Hubble UDF)



Stochastic measurements and deep-field imaging call for different approaches.

- Deep-field imaging:
- Exploration mostly wasted.
- Goal of adaptive strategy:
 - --> quickly converge and *exploit*.



- Stochastic fluctuations:
- Exploration mitigates cosmic variance.



Stochastic measurements and deep-field imaging call for different approaches.

- Deep-field imaging:
- Exploration mostly wasted.
- Goal of adaptive strategy:
 - --> quickly converge and *exploit*.



- Stochastic fluctuations:
- Exploration mitigates cosmic variance.

Goal of adaptive strategy:

--> find ideal patches to *exploit*.



• A preferred method of detection is prolonged integration over a small sky patch.

- A preferred method of detection is prolonged integration over a small sky patch.
- Tradeoff is between *finding lower-foreground patches* and *integrating over them*.

- A preferred method of detection is prolonged integration over a small sky patch.
- Tradeoff is between *finding lower-foreground patches* and *integrating over them*.



Templates for polarized emission from dust (PED) in the Galaxy at 150GHz

(Clark et al. arXiv:1211.6404)





































Outline

Outline

• The Multi-Armed-Bandit Problem
The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

Mitigating CMB B-mode Foregrounds

The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

Mitigating CMB B-mode Foregrounds

MAB Strategies Elsewhere



• The goal:

Facing slots with different odds, maximize winnings.



• The goal:

Facing slots with different odds, maximize winnings.

• With infinite funds, this is easy. You *learn* the odds.



• The goal:

Facing slots with different odds, maximize winnings.

- With infinite funds, this is easy. You *learn* the odds.
- With a finite number of plays, problem is unsolved.



• The goal:

Facing slots with different odds, maximize winnings.

- With infinite funds, this is easy. You *learn* the odds.
- With a finite number of plays, problem is unsolved.
- Heuristics have been developed and compared.





 $\mu^*(a)$

An MAB strategy:



 $\mu^*(a)$

An MAB strategy:

• Expected reward estimates are based on a sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \dots + r_{N_t(a)}}{N_t(a)}$$



An MAB strategy:

• Expected reward estimates are based on a sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \dots + r_{N_t(a)}}{N_t(a)} \qquad \mu_t(a) \underset{N_t(a) \to \infty}{\longrightarrow} \mu^*(a)$$



An MAB strategy:

• Expected reward estimates are based on a sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \dots + r_{N_t(a)}}{N_t(a)} \qquad \mu_t(a) \underset{N_t(a) \to \infty}{\longrightarrow} \mu^*(a)$$

• These estimates are then used in order to choose the action at t + 1.



An MAB strategy:

• Expected reward estimates are based on a sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \dots + r_{N_t(a)}}{N_t(a)} \qquad \mu_t(a) \underset{N_t(a) \to \infty}{\longrightarrow} \mu^*(a)$$

- These estimates are then used in order to choose the action at t + 1.
- Strategies can be compared to one another or to an asymptotic lower bound. (Lai & Robbins, 1985)



The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$



BICEP (larger $f_{\rm sky}$)

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

Never explores --> has linear total regret.



BICEP (larger $f_{\rm sky}$)

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

Never explores --> has linear total regret.



BICEP (larger $f_{\rm sky}$)

• Greedy

$$p_t(a) = 1$$
 if $a = \underset{a'}{\operatorname{argmax}} \mu_t(a')$

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

Never explores --> has linear total regret.



BICEP (larger $f_{\rm sky}$)

• Greedy

$$p_t(a) = 1$$
 if $a = \underset{a'}{\operatorname{argmax}} \mu_t(a')$



POLARBEAR (small f_{sky})

• Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

Never explores --> has linear total regret.



BICEP (larger $f_{\rm sky}$)

• Greedy

$$p_t(a) = 1$$
 if $a = \underset{a'}{\operatorname{argmax}} \mu_t(a')$

Never explores --> has linear total regret.



POLARBEAR (small f_{sky})

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

- ϵ_t greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$
 - --> can achieve logarithmic regret!

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$

--> can achieve logarithmic regret!

 Probability matching (Boltzmann)

$$p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a'} e^{\mu_t(a')/\tau}}$$

(W.R. Thompson, 1933)

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{\alpha}$

--> can achieve logarithmic regret!

 Probability matching (Boltzmann)

$$p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a'} e^{\mu_t(a')/\tau}}$$

(W.R. Thompson, 1933)

Limits: $\tau \to 0$ (greedy) $\tau \to \infty$ (uniformly random)

• ϵ - greedy $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$

--> can achieve logarithmic regret!

 Probability matching (Boltzmann)

$$p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a'} e^{\mu_t(a')/\tau}}$$

(W.R. Thompson, 1933)

Limits: $\tau \to 0$ (greedy) $\tau \to \infty$ (uniformly random)

• Upper Confidence Bound (UCB)

(Auer, Cesa-Bianchi, and Fischer, 2002)

• ϵ - greedy

 $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$

--> can achieve logarithmic regret!

 Probability matching (Boltzmann) $p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a'} e^{\mu_t(a')/\tau}}$

(W.R. Thompson, 1933)

Limits: $\tau \to 0$ (greedy) $\tau \to \infty$ (uniformly random)

• Upper Confidence Bound (UCB) (Auer, Cesa-Bianchi, and Fischer, 2002) define: $\mu^*(a) \le \mu_t(a) + U_t(a)$ and choose: $a_t = \operatorname*{argmax}_a \{\mu_t(a) + U_t(a)\}$

• ϵ - greedy

 $\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \epsilon & \text{Uniformly Random} \end{cases}$

Always explores --> has linear total regret.

• ϵ_t - greedy with a decaying strategy: $\epsilon_t \propto 1/t^{lpha}$

--> can achieve logarithmic regret!

 Probability matching (Boltzmann) $p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a'} e^{\mu_t(a')/\tau}}$

(W.R. Thompson, 1933)

Limits: $\tau \to 0$ (greedy) $\tau \to \infty$ (uniformly random)

• Upper Confidence Bound (UCB) (Auer, Cesa-Bianchi, and Fischer, 2002) define: $\mu^*(a) \le \mu_t(a) + U_t(a)$ and choose: $a_t = \underset{a}{\operatorname{argmax}} \{\mu_t(a) + U_t(a)\}$ for normal distributions: $U_t(a) \propto \sigma_a / \sqrt{N_t(a)}$

Heuristic Solution Algorithms: Simulation



Heuristic Solution Algorithms: Simulation

• MAB Simulation: $\langle \dots \rangle_{1,000}$


Heuristic Solution Algorithms: Simulation

• MAB Simulation: $\langle \dots \rangle_{1,000}$ $n_p = 10; \forall a: \mathcal{N}(\mu^*(a), \sigma_a)$ $n_p = 10; \forall a: \mathcal{N}(\mu^*(a), \sigma_a)$

Heuristic Solution Algorithms: Simulation



Heuristic Solution Algorithms: Simulation



Outline

Outline

The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

Mitigating CMB B-mode Foregrounds











What do we measure?





• PED ``1-sigma'' uncertainty:

$$\sigma_{\ell}^{\widehat{A}_{p}} = \sqrt{\frac{2}{f_{\text{sky}}(2\ell+1)}} \left(A_{p} \tilde{C}_{\ell}^{D} + \alpha C_{\ell}^{L} + f_{\text{sky}} w^{-1}(T) e^{\ell^{2} \sigma_{b}^{2}} \right)$$

• PED ``1-sigma'' uncertainty: foregrounds $\sigma_{\ell}^{\widehat{A}_{p}} = \sqrt{\frac{2}{f_{\rm sky}(2\ell+1)}} \begin{pmatrix} \varphi^{\rm foregrounds} & \varphi^{\rm foregrounds} \\ \left(A_{p}\tilde{C}_{\ell}^{D} + \alpha C_{\ell}^{L} + f_{\rm sky}w^{-1}(T)e^{\ell^{2}\sigma_{b}^{2}}\right)$

• PED ``1-sigma'' uncertainty: foregrounds lensing $\sigma_{\ell}^{\widehat{A}_{p}} = \sqrt{\frac{2}{f_{\rm sky}(2\ell+1)}} \begin{pmatrix} {\rm foregrounds} & {\rm lensing} \\ \downarrow & \downarrow \\ \left(A_{p}\tilde{C}_{\ell}^{D} + \alpha C_{\ell}^{L} + f_{\rm sky}w^{-1}(T)e^{\ell^{2}\sigma_{b}^{2}}\right)$

• PED ``1-sigma'' uncertainty: foregrounds lensing instrumental noise $\sigma_{\ell}^{\widehat{A}_{p}} = \sqrt{\frac{2}{f_{\text{sky}}(2\ell+1)}} \left(A_{p}\tilde{C}_{\ell}^{D} + \alpha C_{\ell}^{L} + f_{\text{sky}}w^{-1}(T)e^{\ell^{2}\sigma_{b}^{2}} \right)$









• In a *pessimistic* scenario - high dust fraction, no de-lensing:

• In a *pessimistic* scenario - high dust fraction, no de-lensing:



• In an *optimistic* scenario - low dust fraction, 80% de-lensing:

• In an optimistic scenario - low dust fraction, 80% de-lensing:



Pessimistic

Optimistic



Optimistic



Outline
Outline

The Multi-Armed-Bandit Problem

Heuristic Solution Algorithms

• Mitigating CMB B-mode Foregrounds

• 21-cm stochastic fluctuations.

• 21-cm stochastic fluctuations.



21-cm stochastic fluctuations.
A 3D-bandit problem.



 \mathcal{V}

Recombination

1100

21-cm stochastic fluctuations.
A 3D-bandit problem.

Radio Interferometer

21-cm stochastic fluctuations.
A 3D-bandit problem.



Radio Interferometer









MAB Elsewhere

MAB Elsewhere

Deep-field imaging:
From HST to JWST?









Optimal Patch



Optimal Patch



Measuring the Luminosity function and calibrating simulations. (with P. Behroozi @CANDELS)

Measuring the Luminosity function and calibrating simulations. (with P. Behroozi @CANDELS)

• Cosmic variance uncertainty plays a huge role.

Measuring the Luminosity function and calibrating simulations. (with P. Behroozi @CANDELS)

• Cosmic variance uncertainty plays a huge role.



Measuring the Luminosity function and calibrating simulations. (with P. Behroozi @CANDELS)

• Cosmic variance uncertainty plays a huge role.



Measuring the Luminosity function and calibrating simulations. (with P. Behroozi @CANDELS)

• Cosmic variance uncertainty plays a huge role.



CANDELS fields and HUDF differ at $z \sim 7,8$ by factors of 3-4.

An efficient adaptive strategy would converge onto the cosmic mean.
(Not useful in a Casino, but may save considerable telescope time)

Thank you!

Ely D. Kovetz University of Texas at Austin ITC Seminar, Dec. 10th, 2013

Thank you!

Ely D. Kovetz University of Texas at Austin ITC Seminar, Dec. 10th, 2013





Thank you!

Ely D. Kovetz University of Texas at Austin ITC Seminar, Dec. 10th, 2013





Discussion: B-mode detection

Discussion: B-mode detection

Discussion: B-mode detection

Recall some assumptions:

• Single frequency. In practice, remain with (non-zero) foreground residuals.

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .
- Limited scenarios. More depth between pessimistic and optimistic.

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .
- Limited scenarios. More depth between pessimistic and optimistic.
- Cost of moving telescope target. This *should* be taken into account.

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .
- Limited scenarios. More depth between pessimistic and optimistic.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:
Recall some assumptions:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .
- Limited scenarios. More depth between pessimistic and optimistic.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

• Methods can be optimized further.

Recall some assumptions:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Specific template (FGPol). Real foregrounds may differ, probably not by much.
- Power-law power spectrum. Exact shape can be calculated from template.
- Gaussian likelihood curves. Full simulation would draw from χ^2 at each ℓ .
- Limited scenarios. More depth between pessimistic and optimistic.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

- Methods can be optimized further.
- Identification of optimal patches will reap future benefits.

Full B-mode simulation results:

Full B-mode simulation results:

• 25%-75% improvement on average in different scenarios.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

• Single frequency. In practice, remain with (non-zero) foreground residuals.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Cost of moving telescope target. This *should* be taken into account.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

• Methods can be optimized further.

Full B-mode simulation results:

- 25%-75% improvement on average in different scenarios.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

- Methods can be optimized further.
- Identification of optimal patches will reap future benefits.