Multiarm Bandits

Ryan T. Moore

2024-08-22

What's Your Strategy?

Suppose 10mins in room of slot machines, huge pile of tokens.

Suppose 10mins in room of slot machines, huge pile of tokens.
Goal: finish w/ ¥¥¥!

Suppose 10mins in room of slot machines, huge pile of tokens.
Goal: finish w/ ¥¥¥!

What would you do? Discuss!

Multiarm Bandits

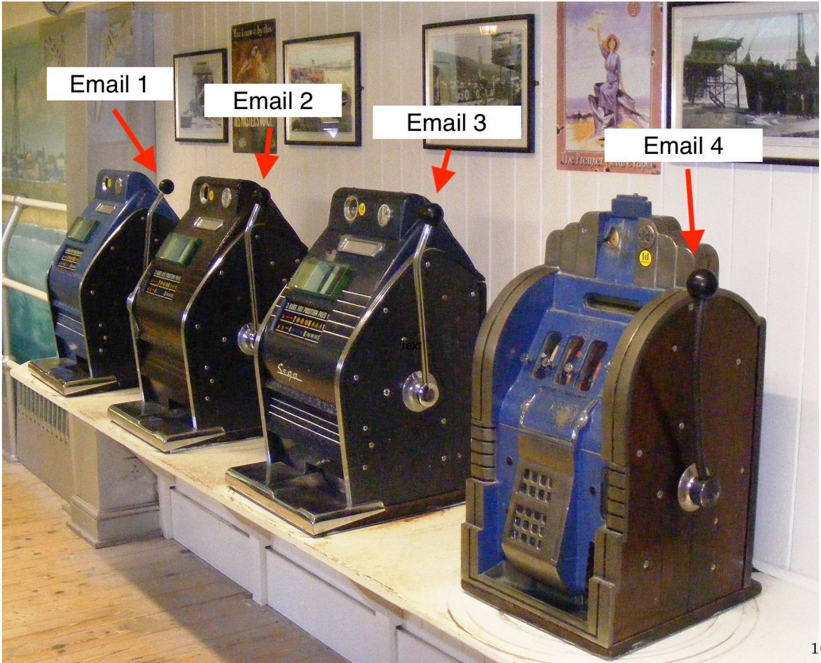
Simultaneously *identify* and *play* machine ("one-armed bandit") with biggest payoff.

Multiarm Bandits

Simultaneously *identify* and *play* machine ("one-armed bandit") with biggest payoff.

Explore all machines. *Exploit* best.

Political Multiarm Bandits



Multiarm Bandits

- Goal: maximize some reward by ID best treatment arm

Multiarm Bandits

- Goal: maximize some reward by ID best treatment arm
- An estimand: not the ATE, but

$$P\left(E(R_{\text{this arm}}) > E(R_{\text{all other arms}})\right)$$

$$P\left(\text{this email's donations} > \text{all other emails' donations}\right)$$

Multiarm Bandits

- Goal: maximize some reward by ID best treatment arm
- An estimand: not the ATE, but

$$P\left(E(R_{\text{this arm}}) > E(R_{\text{all other arms}})\right)$$

$$P\left(\text{this email's donations} > \text{all other emails' donations}\right)$$

Multiarm Bandits

- Goal: maximize some reward by ID best treatment arm
- An estimand: not the ATE, but

$$P\left(E(R_{\text{this arm}}) > E(R_{\text{all other arms}})\right)$$

$$P\left(\text{this email's donations} > \text{all other emails' donations}\right)$$

Situation:

- Testing many arms
- Political/policy goals alongside research goals
- Sniderman's "sequential factorials" all at once

Multiarm Bandits

The "exploration vs exploitation" perspective
Offer-Westort, Coppock, and Green (2021)

Multiarm Bandits

The "exploration vs exploitation" perspective
Offer-Westort, Coppock, and Green (2021)

- "*explore* by obtaining information about the probability of success of each arm so that they can be confident in selecting the best arm or arms."

Multiarm Bandits

The "exploration vs exploitation" perspective
Offer-Westort, Coppock, and Green (2021)

- "*explore* by obtaining information about the probability of success of each arm so that they can be confident in selecting the best arm or arms."
- "*exploit* the best performing arms by allocating large proportions of subjects to them"

Multiarm Bandits

The "exploration vs exploitation" perspective
Offer-Westort, Coppock, and Green (2021)

- "*explore* by obtaining information about the probability of success of each arm so that they can be confident in selecting the best arm or arms."
- "*exploit* the best performing arms by allocating large proportions of subjects to them"
- $\Rightarrow\Leftarrow$

Intro to Adaptive Designs

Adaptive Designs

- Sequentially-blocked designs
  ("covariate adaptive", Moore and Moore (2013))

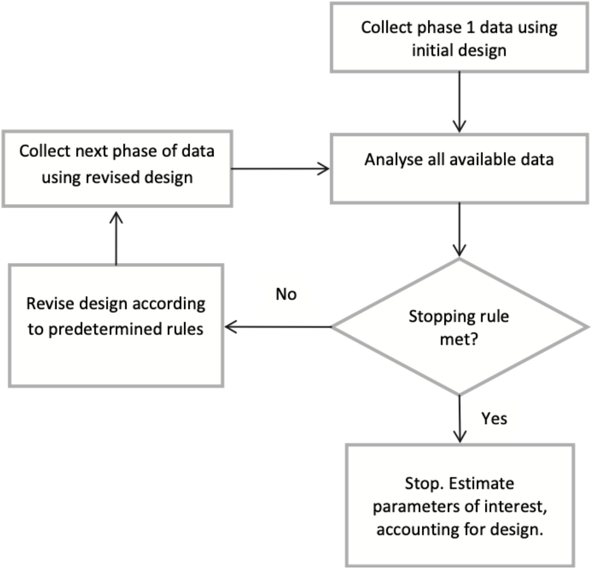
Adaptive Designs

- Sequentially-blocked designs
  ("covariate adaptive", Moore and Moore (2013))
- Multiarm bandits
  ("outcome adaptive")

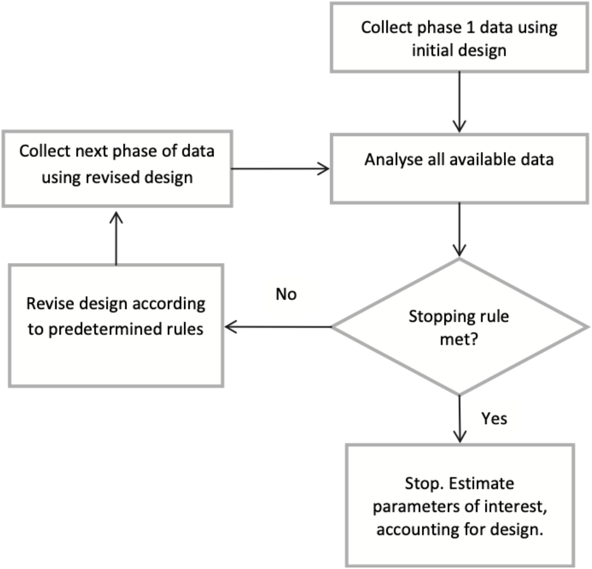
Adaptive Designs

- Sequentially-blocked designs
  ("covariate adaptive", Moore and Moore (2013))
- Multiarm bandits
  ("outcome adaptive")
- (Bayesian) stopping rules
  (strongly adaptive)

Stopping Rules



Stopping Rules



Tourangeau et al. (2017)

Traditional Adaptive Design Motivation

Very different statistical goals:

Traditional Adaptive Design Motivation

Very different statistical goals:

1. Minimize variance
2. Minimize non-response bias (or proxies)
3. Maximize response rates

Tourangeau et al. (2017)

Multiarm Bandit Applications

Offer-Westort, Coppock, and Green (2021)

	Minimum Wage	Right to Work
Question Text	Imagine that the following ballot measure were up for a vote in your state. The measure would: [**ballot measure text**]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]	Imagine that the following ballot measure were up for a vote in your state. The measure would [**amend the State Constitution to**]: [**ballot measure text**]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]
Proposal 1	increase the minimum wage [**from {current}**] to {current + 1} per hour, adjusted annually for inflation, and provide that no more than $3.02 per hour in tip income may be used to offset the minimum wage of employees who regularly receive tips.	prohibit, as a condition of employment, forced membership in a labor organization (union) or forced payments of dues or fees, in full or pro-rata ("fair-share"), to a union. The measure will also make any activity which violates employees' rights provided by the bill illegal and ineffective and allow legal remedies for anyone injured as a result of another person violating or threatening to violate those employees' rights. The measure will not apply to union agreements entered into before the effective date of the measure, unless those agreements are amended or renewed after the effective date of the measure.
Proposal 2	raise the minimum wage [**from {current}**] to {current + 1} per hour effective September 30th, 2021. Each September 30th thereafter, minimum wage shall increase by $1.00 per hour until the minimum wage reaches {current + 5} per hour on September 30th, 2026. From that point forward, future minimum wage increases shall revert to being adjusted annually for inflation starting September 30th, 2027.	The right of persons to work may not be denied or abridged on account of membership or nonmembership in any labor union or labor organization, and all contracts in negation or abrogation of such rights are hereby declared to be invalid, void, and unenforceable.
Proposal 3	Shall the minimum wage for adults over the age of 18 be raised [**from {current}**] to {current + 1} per hour by January 1, 2019?	ban any new employment contract that requires employee to resign from or belong to a union, pay union dues, or make other payment to a union. Required contributions to charity or other third party instead of payments to union are also banned. Employees must authorize payroll deduction to unions. Violations of the section is a misdemeanor.
Proposal 4	raise the minimum wage [**from {current}**] to {current + 1} per hour worked if the employer	No person shall be deprived of life, liberty or property without due process of law. The right of

Offer-Westort, Coppock, and Green (2021)

A sample empirical finding:

Right-to-work proposals tend to be more popular as *ballot measures* than as *constitutional amendments*.

Multiarm Bandit Designs

Three Canonical Designs

- Thompson sampling
- Upper confidence bound
- ϵ-greedy

Thompson Sampling

- Suppose emails A, B, C

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, $\ldots$

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, $\ldots$
- Assign next voter to email, w/ prob. that email is best

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, $\ldots$
- Assign next voter to email, w/ prob. that email is best

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, $\ldots$
- Assign next voter to email, w/ prob. that email is best

Let $P(A \text{ best}) = .6$, $P(B \text{ best}) = .1$, $P(C \text{ best}) = .3$

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, $\ldots$
- Assign next voter to email, w/ prob. that email is best

Let $P(A \text{ best}) = .6$, $P(B \text{ best}) = .1$, $P(C \text{ best}) = .3$

Then, assign next recipient to

- A with prob 0.6
- B with prob 0.1
- C with prob 0.3

Thompson Sampling

- Suppose emails A, B, C
- Let $\theta^A = P(\text{donation})$ from sending A, ...
- Assign next voter to email, w/ prob. that email is best

Let $P(A \text{ best}) = .6$, $P(B \text{ best}) = .1$, $P(C \text{ best}) = .3$

Then, assign next recipient to

- A with prob 0.6
- B with prob 0.1
- C with prob 0.3

- Also, "probability matching"

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

- Assign unit i to treatment arm a^k s.t.

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

- Assign unit i to treatment arm a^k s.t.

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

- Assign unit i to treatment arm a^k s.t.

$$P(\text{assign } a^k \text{ to } i) \quad = \quad P(a^k \text{ is best})$$

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

- Assign unit i to treatment arm a^k s.t.

$$
\begin{aligned}
P(\text{assign } a^k \text{ to } i) &= P(a^k \text{ is best}) \\
P(a_i = a^k) &= P(\theta^k > \theta^1 \text{ and } \theta^k > \theta^2 \text{ and } \ldots \; \theta^k > \theta^K)
\end{aligned}
$$

Thompson Sampling

- Let $k \in \{1, \ldots, K\}$ index arms a
- Let $\theta^k = P(\text{reward})$ from a^k
- (Or, $\theta^k = E(\text{reward})$ from a^k)

- Assign unit i to treatment arm a^k s.t.

$$
\begin{aligned}
P(\text{assign } a^k \text{ to } i) &= P(a^k \text{ is best}) \\
P(a_i = a^k) &= P(\theta^k > \theta^1 \text{ and } \theta^k > \theta^2 \text{ and } \ldots\ \theta^k > \theta^K)
\end{aligned}
$$

- Estimate the θ^k with

$$
\hat{\theta}^k = \text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})
$$

True/False Quiz

1. Multiarm bandits ignore outcomes when assigning treatments.
2. Multiarm bandits try to maximize rewards, not just estimate causal effects.
3. Thompson sampling assigns a unit to the treatments with equal probability.

UCB Sampling

- Let n_k be # units already assigned to a^k

UCB Sampling

- Let n_k be # units already assigned to a^k
- Calculate $\text{UCB}^k \equiv E[\hat{\theta}^k] + \sqrt{\frac{2 \ln N}{n_k}}$ for each a^k

UCB Sampling

- Let n_k be # units already assigned to a^k
- Calculate $\text{UCB}^k \equiv E[\hat{\theta}^k] + \sqrt{\frac{2 \ln N}{n_k}}$ for each a^k
- Assign unit i to treatment a^k iff $\text{UCB}^k > \text{UCB}^1$ and $\text{UCB}^k > \text{UCB}^2$ and $\ldots$ and $\text{UCB}^k > \text{UCB}^K$

UCB Sampling

- Let n_k be # units already assigned to a^k
- Calculate $\text{UCB}^k \equiv E[\hat{\theta}^k] + \sqrt{\frac{2 \ln N}{n_k}}$ for each a^k
- Assign unit i to treatment a^k iff $\text{UCB}^k > \text{UCB}^1$ and $\text{UCB}^k > \text{UCB}^2$ and ... and $\text{UCB}^k > \text{UCB}^K$
- Deterministic

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., $\text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})$

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., Beta(α_0^k + successes, β_0^k + failures)
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., $\text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})$
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., $\text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})$
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., $\text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})$
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

- Can set $\epsilon = \min\left(1, \frac{cK}{d^2 n}\right)$

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., Beta(α_0^k + successes, β_0^k + failures)
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

- Can set $\epsilon = \min\left(1, \frac{cK}{d^2 n}\right)$
- As $n \to \infty$, $P(a^k)$ increases

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., $\text{Beta}(\alpha_0^k + \text{successes}, \beta_0^k + \text{failures})$
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

- Can set $\epsilon = \min\left(1, \frac{cK}{d^2 n}\right)$
- As $n \to \infty$, $P(a^k)$ increases
- Tune with $c > 0$, $0 < d < 1$

ϵ-greedy

- Calculate $\hat{\theta}^k$ from, e.g., Beta(α_0^k + successes, β_0^k + failures)
- Assign unit i to treatment a^k with prob $1 - \epsilon$ if $\hat{\theta}^k$ is largest
- Assign unit i to treatment $a^{\sim k}$ with $\frac{\epsilon}{K-1}$ otherwise

- Can set $\epsilon = \min\left(1, \frac{cK}{d^2 n}\right)$
- As $n \to \infty$, $P(a^k)$ increases
- Tune with $c > 0$, $0 < d < 1$
- ($n = \sum n_k$)

Composite Designs

- 90% assigned via Thompson sampling
- 10% assigned via $a^i \sim \text{Unif}(a^1, a^2, \ldots, a^k)$

Composite Designs

- 90% assigned via Thompson sampling
- 10% assigned via $a^i \sim \text{Unif}(a^1, a^2, \ldots, a^k)$

- Ensuring exploration

Composite Designs

- 90% assigned via Thompson sampling
- 10% assigned via $a^i \sim \text{Unif}(a^1, a^2, \ldots, a^k)$

- Ensuring exploration
- Offer-Westort, Coppock, and Green (2021)'s *control-augmented* algorithm

Composite Designs

- 90% assigned via Thompson sampling
- 10% assigned via $a^i \sim \text{Unif}(a^1, a^2, \ldots, a^k)$

- Ensuring exploration
- Offer-Westort, Coppock, and Green (2021)'s *control-augmented* algorithm
  - Intuition: try to keep $n_{\text{current best}} \approx n_{\text{control}}$

Some Bayesian Background

Law of Total Probability

Decompose $P(A)$ into two components: A happening when B also happens, and A happening when "not B" happens:

$$P(A) = P(A \text{ and } B) + P(A \text{ and } B^C)$$

Law of Total Probability

Decompose $P(A)$ into two components: A happening when B also happens, and A happening when "not B" happens:

$$P(A) = P(A \text{ and } B) + P(A \text{ and } B^C)$$

Similarly, if B_i events (a) are mutually exclusive, and (b) cover the entire sample space, then for B_1, B_2, $\ldots$, B_N,

$$P(A) = \sum_{i=1}^{N} P(A \text{ and } B_i)$$

Bayes' Rule

Joint probability $P(A \text{ and } B)$ is both

$$P(A \text{ and } B) = P(A|B)P(B)$$

and

$$P(A \text{ and } B) = P(B|A)P(A)$$

Bayes' Rule

Joint probability $P(A \text{ and } B)$ is both

$$P(A \text{ and } B) = P(A|B)P(B)$$

and

$$P(A \text{ and } B) = P(B|A)P(A)$$

Set equal, then

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

Bayes' Rule

Joint probability $P(A \text{ and } B)$ is both

$$P(A \text{ and } B) = P(A|B)P(B)$$

and

$$P(A \text{ and } B) = P(B|A)P(A)$$

Set equal, then

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

- B: the data we observe

Bayes' Rule

Joint probability $P(A \text{ and } B)$ is both

$$P(A \text{ and } B) = P(A|B)P(B)$$

and

$$P(A \text{ and } B) = P(B|A)P(A)$$

Set equal, then

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

- B: the data we observe
- $P(A)$: *prior* belief about likely values of A

Bayes' Rule

Joint probability $P(A \text{ and } B)$ is both

$$P(A \text{ and } B) = P(A|B)P(B)$$

and

$$P(A \text{ and } B) = P(B|A)P(A)$$

Set equal, then

$$\begin{aligned} P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

- B: the data we observe
- $P(A)$: *prior* belief about likely values of A
- $P(A|B)$: *posterior* estimate. Includes both our prior belief $P(A)$, but updates using *likelihood* of data, $P(B|A)$

Bayes' Rule

$$\begin{aligned}\text{posterior} &= \frac{\text{likelihood} \cdot \text{prior}}{\text{marginal}} \\ &\propto \text{likelihood} \cdot \text{prior}\end{aligned}$$

General Bayes' Rule

From law of total probability,

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B \text{ and } A) + P(B \text{ and } A^C)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} \end{aligned}$$

General Bayes' Rule

From law of total probability,

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\
&= \frac{P(B|A)P(A)}{P(B \text{ and } A) + P(B \text{ and } A^C)} \\
&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}
\end{aligned}
$$

Extend to other partitions of $P(B)$. If A has 3 types, then,

$$
\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\
&= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)}
\end{aligned}
$$

General Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{N} P(B \text{ and } A_i)}$$

- A: parameters we're interested in
- B: the data we observe
- $P(A)$: *prior* belief about likely values of A
- $P(A|B)$: *posterior* estimate. Includes both our prior belief $P(A)$, but updates using *likelihood* of data, $P(B|A)$

Example

In Boston, 30% of people are conservative, 50% are liberal, and 20% are independent. In the last election, 65% of conservatives, 82% of liberals, and 50% of independents voted. If a person in Boston is selected at random and we learn that she did not vote last election, what is the probability she is a liberal?

Beta-binomial Model

Calculating $\hat{\theta}^k$

Beta-binomial Model

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim \text{Bin}(n, \theta)$

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim \text{Bin}(n, \theta)$
- Then, the posterior is

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim \text{Bin}(n, \theta)$
- Then, the posterior is

Beta-binomial Model

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim \text{Bin}(n, \theta)$
- Then, the posterior is

$$\begin{aligned} posterior &\propto \text{Beta}(\alpha, \beta) \cdot \text{Bin}(n, \theta) \\ &\propto \text{Beta}(\alpha + s, \beta + (s - n)) \end{aligned}$$

Beta-binomial Model

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim \text{Bin}(n, \theta)$
- Then, the posterior is

$$\begin{aligned} \textit{posterior} &\propto \text{Beta}(\alpha, \beta) \cdot \text{Bin}(n, \theta) \\ &\propto \text{Beta}(\alpha + s, \beta + (s - n)) \end{aligned}$$

- What are these distributions?

Beta Distribution

- $x \in [0, 1]$ – a model for probability!

Beta Distribution

- $x \in [0, 1]$ – a model for probability!
- $X \sim \mathrm{Beta}(\alpha, \beta)$

Beta Distribution

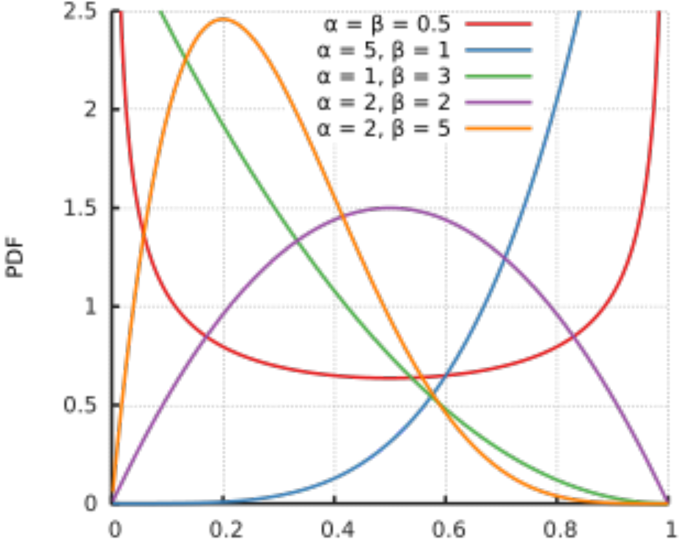
- $x \in [0, 1]$ – a model for probability!
- $X \sim \mathrm{Beta}(\alpha, \beta)$
- $\alpha, \beta > 0$, $\alpha - 1$ successes, $\beta - 1$ failures

Beta Distribution

- $x \in [0, 1]$ – a model for probability!
- $X \sim \text{Beta}(\alpha, \beta)$
- $\alpha, \beta > 0$, $\alpha - 1$ successes, $\beta - 1$ failures
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

Beta Distribution

- $x \in [0, 1]$ – a model for probability!
- $X \sim \text{Beta}(\alpha, \beta)$
- $\alpha, \beta > 0$, $\alpha - 1$ successes, $\beta - 1$ failures
- $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
- $\text{Beta}(1, 1) \sim \text{Unif}(0, 1)$



(Wikipedia, Feb 2019)

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?'', $k \in \{0, \ldots, n\}$

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?'', $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?'', $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?", $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- Sum of n Bernoullis

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?' ', $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- Sum of n Bernoullis
- Political examples:

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?' ', $k \in \{0, \dots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- Sum of n Bernoullis
- Political examples:
  - prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$ `dbinom(3, 6, prob = .3)` $\approx .19$

Binomial Distribution

- n independent, identically distributed (iid) trials, binary outcome $0, 1$.
- "Prob of k successes in n trials?'', $k \in \{0, \ldots, n\}$
- $X \sim Bin(n, p)$
- $p(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
- Sum of n Bernoullis
- Political examples:
  - prob 3 of 6 opposing Senators support an amendment: $p(X = 3|n = 6, p = .3) =$ `dbinom(3, 6, prob = .3)` $\approx .19$
  - prob ≥ 3 of 6 opposing Senators support an amendment: $p(X \geq 3|n = 6, p = .3) =$ `1 - pbinom(2, 6, prob = .3)` $=$ `pbinom(2,6,prob=.3,lower.tail=FALSE)` $\approx .26$

Calculating $\hat{\theta}^k$

- Let our prior for $\theta \sim \text{Beta}(\alpha, \beta)$
- Let the DGP be $Y \sim Bin(n, \theta)$
- Let s be number of successes
- Then, the posterior is ...

$$\text{posterior} \quad \propto \quad \text{prior} \cdot \text{likelihood}$$

$$\begin{aligned}\text{posterior} &\propto \text{prior} \cdot \text{likelihood} \\ P(\theta|Y) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{s}\theta^{s}(1-\theta)^{n-s}\end{aligned}$$

$$
\begin{aligned}
\text{posterior} &\propto \text{prior} \cdot \text{likelihood} \\
P(\theta|Y) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{s}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \frac{n!}{s!(n-s)!}\theta^{s}(1-\theta)^{n-s}
\end{aligned}
$$

$$
\begin{aligned}
\text{posterior} &\propto \text{prior} \cdot \text{likelihood} \\
P(\theta|Y) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{s}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \frac{n!}{s!(n-s)!}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)} \cdot \\
&\qquad \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^{s}(1-\theta)^{n-s}
\end{aligned}
$$

$$
\begin{aligned}
\text{posterior} &\propto \text{prior} \cdot \text{likelihood} \\
P(\theta|Y) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{s}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \frac{n!}{s!(n-s)!}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)} \cdot \\
&\quad\ \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)} \\
&\quad\ \theta^{\alpha+s-1}(1-\theta)^{\beta+(n-s)-1}
\end{aligned}
$$

$$
\begin{aligned}
\text{posterior} &\propto \text{prior} \cdot \text{likelihood} \\
P(\theta|Y) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{s}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \frac{n!}{s!(n-s)!}\theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)} \cdot \\
&\qquad \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^{s}(1-\theta)^{n-s} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)} \\
&\qquad \theta^{\alpha+s-1}(1-\theta)^{\beta+(n-s)-1} \\
&\propto \text{Beta}(\alpha+s, \beta+(s-n))
\end{aligned}
$$

Beta-binomial Model

- For binomial (sum of $0/1$) outcome data, use Beta prior.

Beta-binomial Model

- For binomial (sum of 0/1) outcome data, use Beta prior.
- Set α, β to be prior successes, failures

Beta-binomial Model

- For binomial (sum of 0/1) outcome data, use Beta prior.
- Set α, β to be prior successes, failures
- Posterior (after data) distribution of θ is Beta

Beta-binomial Model

- For binomial (sum of 0/1) outcome data, use Beta prior.
- Set α, β to be prior successes, failures
- Posterior (after data) distribution of θ is Beta
- Calculate $P(\theta^k > \theta^1)$, etc.

Offer-Westort, Coppock, and Green (2021)

- Simulations we'll replicate below

Offer-Westort, Coppock, and Green (2021)

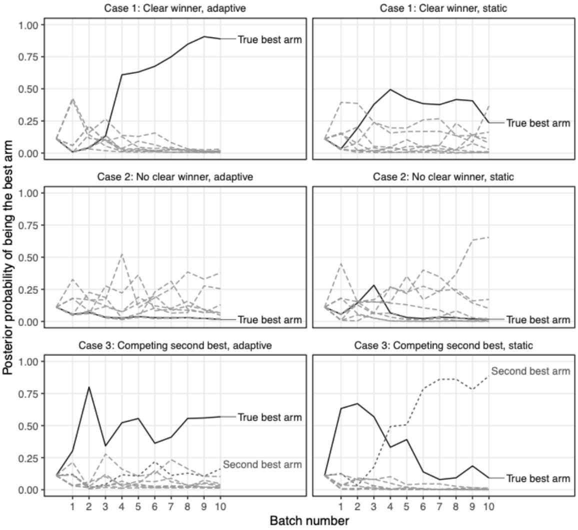
- Simulations we'll replicate below
- Experiment: Finding "best" arguments for ballot proposition elections (minimum wage, right-to-work proposals)

Offer-Westort, Coppock, and Green (2021)

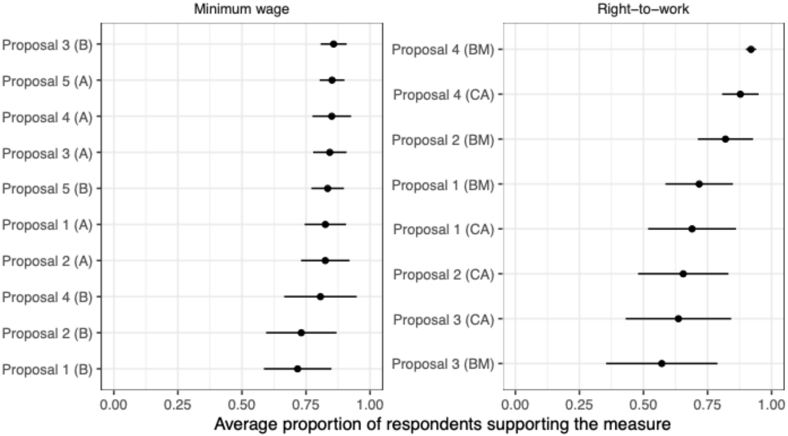
- Simulations we'll replicate below
- Experiment: Finding "best" arguments for ballot proposition elections (minimum wage, right-to-work proposals)
- Experiment: Finding "best" wording for ballot propositions in campaign finance

Offer-Westort, Coppock, and Green (2021), Simulations

Figure 1: Posterior Probabilities Over Time



Offer-Westort, Coppock, and Green (2021), Arguments



Group means are unweighted. “A” versions of the minimum wage proposals include the current minimum wage and “B” versions do not. “CA” versions of the right-to-work proposals are describes as “constitutional amendments” and “BM” versions are not.

Offer-Westort, Coppock, and Green (2021), Wording

Experiment: Finding “best” wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles

Offer-Westort, Coppock, and Green (2021), Wording

Experiment: Finding "best" wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles
- still too big to test them all w/ 1000 participants

Offer-Westort, Coppock, and Green (2021), Wording

Experiment: Finding "best" wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles
- still too big to test them all w/ 1000 participants
- model-assisted, but still adaptive

Offer-Westort, Coppock, and Green (2021), Wording

Experiment: Finding "best" wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles
- still too big to test them all w/ 1000 participants
- model-assisted, but still adaptive
- *less* precision from adaptive than static conjoint

Offer-Westort, Coppock, and Green (2021), Wording

Experiment: Finding "best" wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles
- still too big to test them all w/ 1000 participants
- model-assisted, but still adaptive
- *less* precision from adaptive than static conjoint

Offer-Westort, Coppock, and Green (2021), Wording

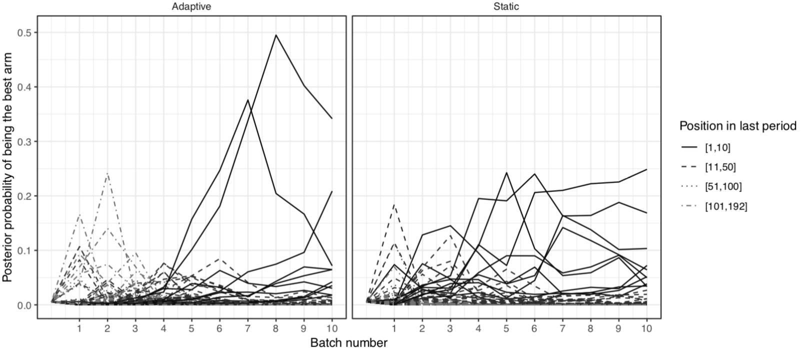
Experiment: Finding "best" wording for ballot propositions in campaign finance

- small conjoint $\rightsquigarrow$ 192 profiles
- still too big to test them all w/ 1000 participants
- model-assisted, but still adaptive
- *less* precision from adaptive than static conjoint

$$SE(\widehat{ATE}) = \sqrt{\frac{1}{N-1}\left[\frac{m\text{Var}(Y_i(0))}{N-m} + \frac{(N-m)\text{Var}(Y_i(1))}{m} + 2\text{Cov}(Y_i(0), Y_i(1))\right]}$$

- But ...

. . . very best most likely to be best



Kuleshov and Precup (2014)

Meta-bandit: pick the best *algorithm* to pick the best arm.

Kuleshov and Precup (2014)

Meta-bandit: pick the best *algorithm* to pick the best arm.

Parameters

- number of treatment conditions
  (2, 5, 10, 50)
- Var(Y)
  ($\sigma \in \{0.01, 0.1, 1\}$)
- reward distribution
  (normal, triangular, uniform, inverse Gaussian, Gumbel)

6 Algorithms Tested

- ϵ-greedy
- Boltzmann exploration
- Pursuit bandits
- Reinforcement comparison
- UCB
- UCB1-Tuned

6 Algorithms Tested

- ϵ-greedy
- Boltzmann exploration
- Pursuit bandits
- Reinforcement comparison
- UCB
- UCB1-Tuned

Roughly,

- More variance in Y: more deterministic UCBs

6 Algorithms Tested

- ϵ-greedy
- Boltzmann exploration
- Pursuit bandits
- Reinforcement comparison
- UCB
- UCB1-Tuned

Roughly,

- More variance in Y: more deterministic UCBs
- Many arms: ϵ-greedy, softmax

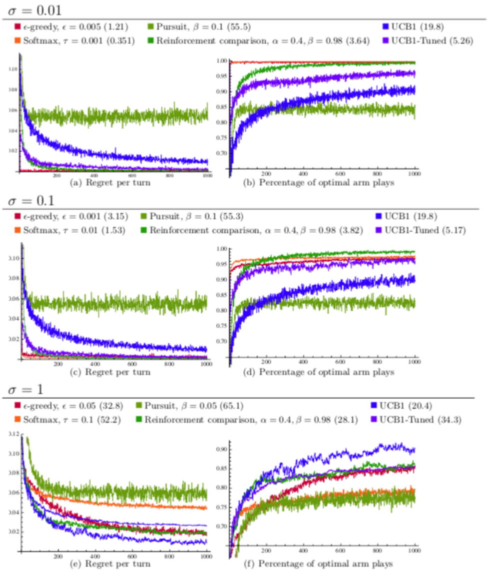


Figure 1: Empirical Results for 2 arms, with different values of the variance

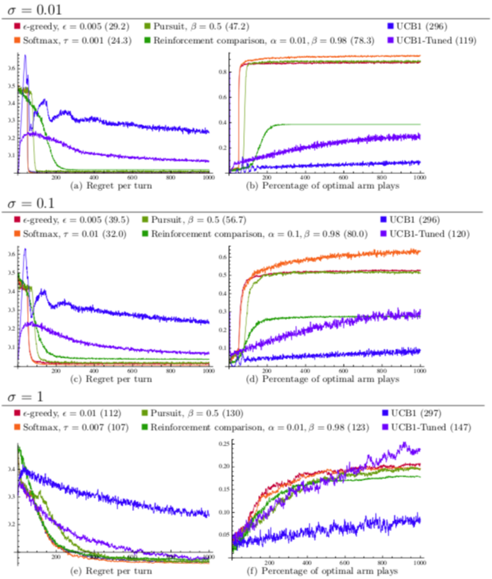


Figure 4: Empirical Results for 50 arms, with different values of the variance

Boltzmann Softmax

- At one extreme, pure greedy algorithm
- At other, uniform choice over a
- Between, like Thompson
  (but prob is proportional, not nec exactly posterior)
- ("temperature" parameter like simulating annealing MCMC exploration)

Jang and Moore (2020)

Nonstationary contextual bandits

Jang and Moore (2020)

Nonstationary contextual bandits

- “nonstationary”: the underlying political world is changing

Jang and Moore (2020)

Nonstationary contextual bandits

- "nonstationary": the underlying political world is changing
- "contextual": heterogeneous treatment effects (e.g., one email for moderates, another for ideological extremes)

Jang and Moore (2020)

Nonstationary contextual bandits

- "nonstationary": the underlying political world is changing
- "contextual": heterogeneous treatment effects (e.g., one email for moderates, another for ideological extremes)

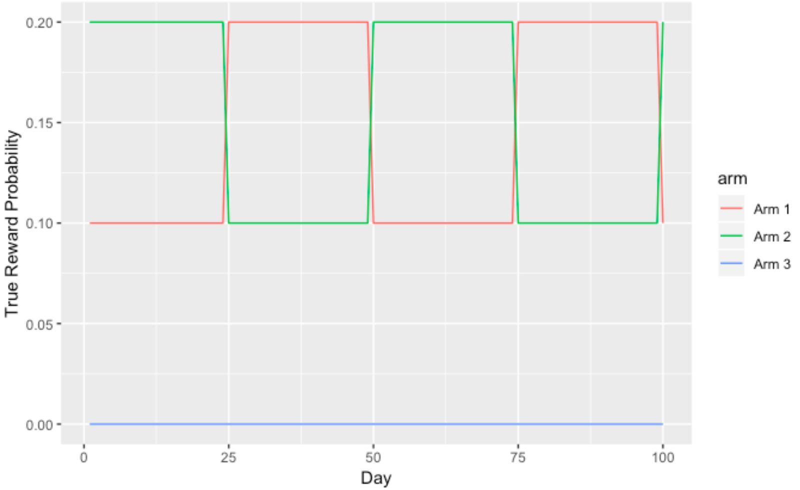
Jang and Moore (2020)

Nonstationary contextual bandits

- "nonstationary": the underlying political world is changing
- "contextual": heterogeneous treatment effects (e.g., one email for moderates, another for ideological extremes)

(Lit has *great* names: sleeping, adversarial, ...)

Political Environments



Jang and Moore (2020)

- Nonstationary contextual bandits superior (outperform stationary, noncontextual; OK if enviro is stationary, etc.)

Jang and Moore (2020)

- Nonstationary contextual bandits superior (outperform stationary, noncontextual; OK if enviro is stationary, etc.)
- "Discounting" old info; "detecting" changes in reward probabilities. Adapting those strategies.

Jang and Moore (2020)

- Nonstationary contextual bandits superior (outperform stationary, noncontextual; OK if enviro is stationary, etc.)
- "Discounting" old info; "detecting" changes in reward probabilities. Adapting those strategies.
- Causal inference intact (despite sample sizes, tr probs, etc.)

Implementation

Implementation

See `code/03-bandits.R`.

Implementation

```r
library(bandit)
library(tidyverse)

set.seed(590646161)
```

Implementation

Clear two-arm trial:

```r
successes <- c(50, 90)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Implementation

Clear two-arm trial:

```r
successes <- c(50, 90)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

Implementation

Clear two-arm trial:

```r
successes <- c(50, 90)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

```
## [1] 1.287403e-10 1.000000e+00
```

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 51)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 51)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 51)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

```
## [1] 0.4440664 0.5559336
```

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 56)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 56)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

Implementation

Competitive two-arm trial:

```r
successes <- c(50, 56)
n <- c(100, 100)

best_binomial_bandit(successes, n)
```

Guess posterior probabilities of being best?

```
## [1] 0.1988609 0.8011391
```

Implementation

Clear five-arm trial:

```r
successes <- c(20, 30, 40, 50, 60)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Implementation

Clear five-arm trial:

```r
successes <- c(20, 30, 40, 50, 60)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Guess posterior probabilities of being best?

Implementation

Clear five-arm trial:

```r
successes <- c(20, 30, 40, 50, 60)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Guess posterior probabilities of being best?

```
## [1] 0.000 0.000 0.001 0.078 0.920
```

Implementation

Competitive five-arm trial:

```r
successes <- c(20, 30, 40, 50, 52)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Implementation

Competitive five-arm trial:

```r
successes <- c(20, 30, 40, 50, 52)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Guess posterior probabilities of being best?

Implementation

Competitive five-arm trial:

```r
successes <- c(20, 30, 40, 50, 52)
n <- c(100, 100, 100, 100, 100)

best_binomial_bandit(successes, n) |> round(3)
```

Guess posterior probabilities of being best?

```
## [1] 0.000 0.000 0.015 0.382 0.602
```

Simulating Binomial Bandits

Three arms:

```r
p_success <- c(0.2, 0.25, 0.3)
n_waves <- 4
n_per_wave <- 20
```

Wave 1: Uniform draw over 3 arms

```r
wave1_arms <- sample(p_success,
                     size = n_per_wave,
                     replace = TRUE)
wave1_arms
```

```
##  [1] 0.25 0.30 0.25 0.20 0.25 0.20 0.25 0.30 0.30 0.30
## [16] 0.25 0.25 0.30 0.25 0.30
```

Simulating Binomial Bandits

```r
table(wave1_arms)
```

```
## wave1_arms
##  0.2 0.25  0.3
##    4    7    9
```

Draw wave 1 outcomes:

```r
wave1_outcome <- rbinom(n_per_wave, 1, prob = wave1_arms)
wave1_outcome
```

```
##  [1] 0 0 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 1 0 1
```

```r
df_wave1 <- tibble(wave1_arms, wave1_outcome)
```

Simulating Binomial Bandits

```r
table_wave1 <- table(df_wave1)
table_wave1
```

```
##            wave1_outcome
## wave1_arms 0 1
##       0.2  2 2
##       0.25 6 1
##       0.3  6 3
```

Simulating Binomial Bandits

Posterior probabilities of being best:

```
successes <- table_wave1[, "1"]
n <- rowSums(table_wave1)

posterior_prob_best <- best_binomial_bandit(successes, n)
posterior_prob_best
```

```
## [1] 0.67814377 0.06351048 0.25834575
```

Simulating Binomial Bandits

Wave 2: Thompson sampling

```r
wave2_arms <- sample(p_success, size = n_per_wave,
                     prob = posterior_prob_best,
                     replace = TRUE)

table(wave2_arms)
```

```
## wave2_arms
## 0.2 0.3
##  17   3
```

```r
wave2_outcome <- rbinom(n_per_wave, 1, prob = wave2_arms)

df_wave2 <- tibble(wave2_arms, wave2_outcome)

table_wave2 <- table(df_wave2)
```

Simulating Binomial Bandits

```
table_wave2
```

```
##             wave2_outcome
## wave2_arms  0  1
##        0.2 15  2
##        0.3  3  0
```

Simulating Binomial Bandits

```
table_wave2
```

```
##             wave2_outcome
## wave2_arms  0  1
##        0.2 15  2
##        0.3  3  0
```

```
successes <- table_wave2[, "1"]
n <- rowSums(table_wave2)

best_binomial_bandit(successes, n)
```

```
## [1] 0.4701299 0.5298701
```

Simulating Binomial Bandits

```r
table_wave2
```

```
##            wave2_outcome
## wave2_arms  0  1
##        0.2 15  2
##        0.3  3  0
```

```r
successes <- table_wave2[, "1"]
n <- rowSums(table_wave2)

best_binomial_bandit(successes, n)
```

```
## [1] 0.4701299 0.5298701
```

(Note: update needed to make *cumulative*!)

Simulating Binomial Bandits

Instead of doing this manually, write a loop/iteration ...

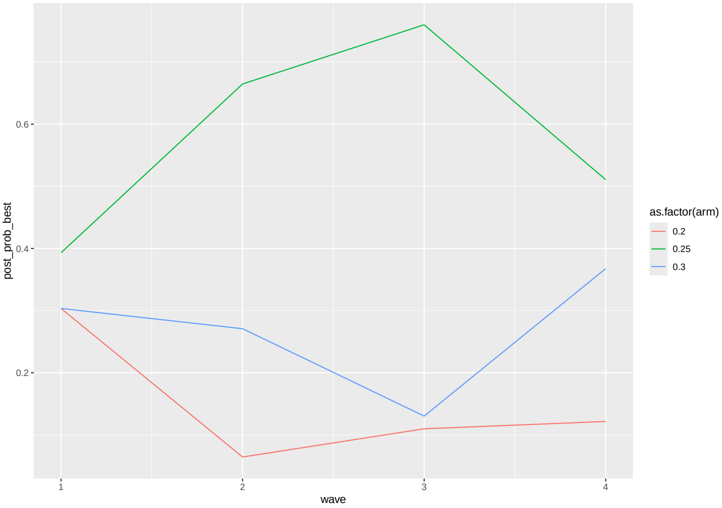
Implement:

```
my_b <- my_ts_bandit(p_success = c(.2, .25, .3),
                     n_waves = 4,
                     n_per_wave = 20)
```

```
## Wave 1 is assigned!
## Wave 2
## Wave 3
## Wave 4
```

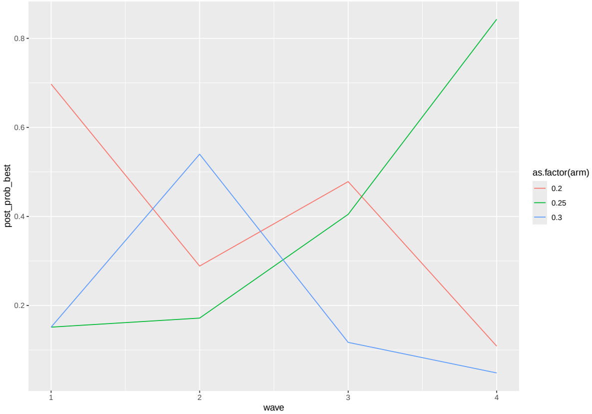
Simulating Binomial Bandits

```
ggplot(my_b, aes(wave, post_prob_best)) +
  geom_line(aes(color = as.factor(arm)))
```



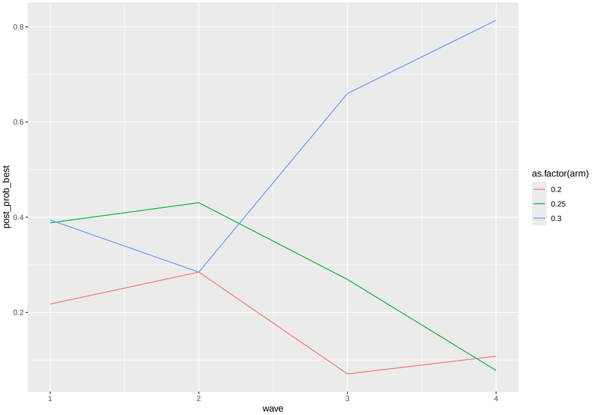
Simulating Binomial Bandits, Take 2

```
my_b <- my_ts_bandit(p_success = c(.2, .25, .3),
               n_waves = 4, n_per_wave = 20, verbose = FALSE)

ggplot(my_b, aes(wave, post_prob_best)) +
  geom_line(aes(color = as.factor(arm)))
```



Simulating Binomial Bandits, Take 3

```
my_b <- my_ts_bandit(p_success = c(.2, .25, .3),
             n_waves = 4, n_per_wave = 20, verbose = FALSE)

ggplot(my_b, aes(wave, post_prob_best)) +
  geom_line(aes(color = as.factor(arm)))
```



Next:
Mediation, Interference,
Transparency, Replication,
Designing Studies?

Jang, Austin, and Ryan T. Moore. 2020. "Maximizing Elusive Rewards: Multiarm Bandits in Dynamic Environments." *Manuscript*.

Kuleshov, Volodymyr, and Doina Precup. 2014. "Algorithms for Multi-Armed Bandit Problems." *CoRR* abs/1402.6028. http://arxiv.org/abs/1402.6028.

Moore, Ryan T., and Sally A. Moore. 2013. "Blocking for Sequential Political Experiments." *Political Analysis* 21 (4): 507–23.

Offer-Westort, Molly, Alexander Coppock, and Donald P. Green. 2021. "Adaptive Experimental Design: Prospects and Applications in Political Science." *American Journal of Political Science* 65 (4): 826–44. https://alexandercoppock.com/offer-westort_coppock_green_2021.pdf.

Tourangeau, Roger, J. Michael Brick, Sharon Lohr, and Jane Li. 2017. "Adaptive and Responsive Survey Designs: A Review and Assessment." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180 (1): 203–23.