When Means Bound Variances: Concentration for Recursively Determined Random Values

Mike Steele [1]

Analysis of Algorithms
Menorca, Spain

May, 2013

[1] Including joint work with Noah Gans and Alessandro Arlotto

"Just once in a while I'd like to see someone begin a talk without a PLAN"

—

Luc Devroye, Barbados, 2013

After starting with a quote — time for a *“heads up”*?

1. Quick Look at Variance Bounds in the “Early Days”
   - The Euclidean TSP: Two Probability Models
   - A Variance Bound for Many Seasons
   - Modern Technology — a Gap and a Challenge
2. A Challenge for Bounders of Variances
   - Leading Example: The Sequential Knapsack Problem
   - MDPs: A General Framework — Served with Alphabet Soup
   - Three Notable Properties
   - Main Result: Variance Bound for a General Class of MDPs
   - Proof sketch
   - Concrete Conjecture in Simplest Context
3. “Take Aways”

Not a Plan — Just Suggestions

1. Quick Look at Variance Bounds in the "Early Days"
   - The Euclidean TSP: Two Probability Models
   - A Variance Bound for Many Seasons
   - Modern Technology — a Gap and a Challenge
2. A Challenge for Bounders of Variances
   - Leading Example: The Sequential Knapsack Problem
   - MDPs: A General Framework — Served with Alphabet Soup
   - Three Notable Properties
   - Main Result: Variance Bound for a General Class of MDPs
   - Proof sketch
   - Concrete Conjecture in Simplest Context
3. "Take Aways"

First imagine n Points in the plane and the shortest path ...

First imagine n Points in the plane and the shortest path ...

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.

First imagine n Points in the plane and the shortest path ...

- I’d draw a picture but I am “Beamer challenged”
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model
- This matters because of Karp's Partitioning Algorithm for the TSP

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model
- This matters because of Karp's Partitioning Algorithm for the TSP
  - Based on the BHH Theorem

$$\lim_{n\to\infty} \frac{L_n}{\sqrt{n}} = \beta_{\mathrm{BHH}} \quad \text{with probability one.}$$

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model
- This matters because of Karp's Partitioning Algorithm for the TSP
  - Based on the BHH Theorem

$$\lim_{n\to\infty} \frac{L_n}{\sqrt{n}} = \beta_{\mathrm{BHH}} \quad \text{with probability one.}$$

  - But BHH is NOT good enough for honest "almost sure" Partitioning Algorithms

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model
- This matters because of Karp's Partitioning Algorithm for the TSP
  - Based on the BHH Theorem

$$\lim_{n\to\infty} \frac{L_n}{\sqrt{n}} = \beta_{\mathrm{BHH}} \quad \text{with probability one.}$$

  - But BHH is NOT good enough for honest "almost sure" Partitioning Algorithms
- Karp and S. (1985) sorted this out: Once needs complete convergence (or better)

First imagine n Points in the plane and the shortest path ...

- I'd draw a picture but I am "Beamer challenged"
- Now, more precisely *imagine how the points arrive*. There are two ways.
  - The Incrementing model
  - The Independent model
- This matters because of Karp's Partitioning Algorithm for the TSP
  - Based on the BHH Theorem

$$\lim_{n\to\infty} \frac{L_n}{\sqrt{n}} = \beta_{\mathrm{BHH}} \quad \text{with probability one.}$$

  - But BHH is NOT good enough for honest "almost sure" Partitioning Algorithms
- Karp and S. (1985) sorted this out: Once needs complete convergence (or better)
- The tool of the time was a Jackknife bound on the variance.

Familiar Now but Still "News You can Use"

Familiar Now but Still "News You can Use"

- There is a VERY GENERAL trick for bounding

$$\mathrm{Var}F(X_1, X_2, \ldots, X_n)$$

Familiar Now but Still "News You can Use"

- There is a VERY GENERAL trick for bounding

$$\mathrm{Var} F(X_1, X_2, ..., X_n)$$

- Steele (1981), generalizing Efron and Stein (1980), showed it is bounded by

$$\frac{1}{2} \sum_{1 \leq i \leq n} E(F(X_1, X_2, ..., X_i, ..., X_n) - F(X_1, X_2, ..., \hat{X}_i, ..., X_n))^2 \quad \heartsuit$$

Familiar Now but Still "News You can Use"

- There is a VERY GENERAL trick for bounding

$$\mathrm{Var} F(X_1, X_2, ..., X_n)$$

- Steele (1981), generalizing Efron and Stein (1980), showed it is bounded by

$$\frac{1}{2} \sum_{1 \le i \le n} E(F(X_1, X_2, ..., X_i, ..., X_n) - F(X_1, X_2, ..., \hat{X}_i, ..., X_n))^2 \quad \heartsuit$$

- For the TSP (and MST, etc.) this can be used to prove that (in $d = 2$) there is a constant C such that for

$$\mathrm{Var} L_n \le C \qquad \text{for all } n$$

Familiar Now but Still "News You can Use"

- There is a VERY GENERAL trick for bounding

$$\mathrm{Var} F(X_1, X_2, ..., X_n)$$

- Steele (1981), generalizing Efron and Stein (1980), showed it is bounded by

$$\frac{1}{2} \sum_{1 \le i \le n} E(F(X_1, X_2, ..., X_i, ..., X_n) - F(X_1, X_2, ..., \hat{X}_i, ..., X_n))^2 \quad \heartsuit$$

- For the TSP (and MST, etc.) this can be used to prove that (in $d = 2$) there is a constant C such that for

$$\mathrm{Var} L_n \le C \qquad \text{for all } n$$

- Even now this may seem surprising. Here, and in many other cases, it gives an very pleasing path to the desired strong laws.

Modern Technology — a Gap and a Challenge

Modern Technology — a Gap and a Challenge

- After intermediate results by several authors, Talagrand proved with his convex distance inequality that the TSP and MST in $d = 2$ have Gaussian tail bounds:

$$P(|L_n - EL_n| \geq x) \leq Ae^{-Bx^2}.$$

Modern Technology — a Gap and a Challenge

- After intermediate results by several authors, Talagrand proved with his convex distance inequality that the TSP and MST in $d = 2$ have Gaussian tail bounds:

$$P(|L_n - EL_n| \geq x) \leq Ae^{-Bx^2}.$$

- The proof of this inequality can be somewhat simplified by using the Spacefilling Curve Heuristic to get a "certificate" for the size of Talangrand's distance (see PTCO for the trick).

Modern Technology — a Gap and a Challenge

- After intermediate results by several authors, Talagrand proved with his convex distance inequality that the TSP and MST in $d = 2$ have Gaussian tail bounds:

$$P(|L_n - EL_n| \geq x) \leq Ae^{-Bx^2}.$$

- The proof of this inequality can be somewhat simplified by using the Spacefilling Curve Heuristic to get a "certificate" for the size of Talangrand's distance (see PTCO for the trick).
- The problem for $d > 2$ seems to have dropped off the radar. The basic variance bound gives

$$\text{Var}L_n \leq Cn^{(d-2)/d} \qquad \text{for all } n;$$

Modern Technology — a Gap and a Challenge

- After intermediate results by several authors, Talagrand proved with his convex distance inequality that the TSP and MST in $d = 2$ have Gaussian tail bounds:

$$P(|L_n - EL_n| \geq x) \leq Ae^{-Bx^2}.$$

- The proof of this inequality can be somewhat simplified by using the Spacefilling Curve Heuristic to get a "certificate" for the size of Talangrand's distance (see PTCO for the trick).
- The problem for $d > 2$ seems to have dropped off the radar. The basic variance bound gives

$$\mathrm{Var} L_n \leq Cn^{(d-2)/d} \qquad \text{for all } n;$$

- More is known, but the full truth remains to be sorted out.

Modern Technology — a Gap and a Challenge

- After intermediate results by several authors, Talagrand proved with his convex distance inequality that the TSP and MST in $d = 2$ have Gaussian tail bounds:

$$P(|L_n - EL_n| \geq x) \leq Ae^{-Bx^2}.$$

- The proof of this inequality can be somewhat simplified by using the Spacefilling Curve Heuristic to get a "certificate" for the size of Talangrand's distance (see PTCO for the trick).
- The problem for $d > 2$ seems to have dropped off the radar. The basic variance bound gives

$$\mathrm{Var} L_n \leq Cn^{(d-2)/d} \qquad \text{for all } n;$$

- More is known, but the full truth remains to be sorted out.
- The text of Boucheron, Lugosi, and Massart develops the bound ♡ in remarkably powerful ways.

Not a Plan — Just Suggestions

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: *include/exclude* Y_t, $1 \leq t \leq n$

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: *include/exclude* Y_t, $1 \leq t \leq n$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max \left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq c \right\},$$

τ_i, the index of the ith item included must be a stopping time

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: *include/exclude* Y_t, $1 \leq t \leq n$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq c \right\},$$

  τ_i, the index of the ith item included must be a stopping time
- Objective: $\sup_\pi \mathbb{E}\left[R_n(\pi)\right]$

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: *include/exclude* Y_t, $1 \le t \le n$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{k : \sum_{i=1}^{k} Y_{\tau_i} \le c\right\},$$

  τ_i, the index of the ith item included must be a stopping time
- Objective: $\sup_\pi \mathbb{E}\left[R_n(\pi)\right]$
- π_n^*: optimal Markov deterministic policy

How to Pack a Knapsack Sequentially and Optimally — a problem going back at least to Coffman et al. (1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: *include/exclude* Y_t, $1 \leq t \leq n$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{k : \sum_{i=1}^{k} Y_{\tau_i} \leq c\right\},$$

  τ_i, the index of the ith item included must be a stopping time
- Objective: $\sup_\pi \mathbb{E}\left[R_n(\pi)\right]$
- π_n^*: optimal Markov deterministic policy

- La pregunta de hoy:

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \ \mathbf{?}$$

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\Big(\ \mathcal{X} \ , \ \mathcal{Y} \ , \ \mathcal{A} \ , \ f \ , \ r \ , \ n \ \Big)$$

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\Big(\mathcal{X} , \mathcal{Y} , \mathcal{A} , f , r , n \Big)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\left(\mathcal{X} , \mathcal{Y} , \mathcal{A} , f , r , n \right)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \dots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\Big(\mathcal{X} \ , \ \mathcal{Y} \ , \ \mathcal{A} \ , \ f \ , \ r \ , \ n \Big)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \ldots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented
- Action space: $\mathcal{A}(t, x, y) \subseteq \mathcal{A}$ is the set of admissible actions for (x, y) at t
  - Knapsack example: "select"; "do not select"

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\Big(\mathcal{X} , \mathcal{Y} , \mathcal{A} , f , r , n \Big)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \ldots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented
- Action space: $\mathcal{A}(t, x, y) \subseteq \mathcal{A}$ is the set of admissible actions for (x, y) at t
  - Knapsack example: "select"; "do not select"
- State transition function: $f(t, x, y, a)$ state that one reaches for $a \in \mathcal{A}(t, x, y)$
  - Knapsack example: $f(t, x, y, \text{select}) = \text{x} - \text{y}$; $f(t, x, y, \text{do not select}) = \text{x}$

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\left(\mathcal{X} \, , \, \mathcal{Y} \, , \, \mathcal{A} \, , \, f \, , \, r \, , \, n \right)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \dots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented
- Action space: $\mathcal{A}(t, x, y) \subseteq \mathcal{A}$ is the set of admissible actions for (x, y) at t
  - Knapsack example: "select"; "do not select"
- State transition function: $f(t, x, y, a)$ state that one reaches for $a \in \mathcal{A}(t, x, y)$
  - Knapsack example: $f(t, x, y, \mathrm{select}) = \mathrm{x} - \mathrm{y}$; $f(t, x, y, \mathrm{do\,not\,select}) = \mathrm{x}$
- Reward function: $r(t, x, y, a)$ reward for taking action a at time t when at (x, y)
  - Knapsack example: $f(t, x, y, \mathrm{select}) = 1$; $f(t, x, y, \mathrm{do\,not\,select}) = 0$

MDPs: The Alphabet Soup Illustrated by the Knapsack Problem

$$\left(\mathcal{X} \; , \; \mathcal{Y} \; , \; \mathcal{A} \; , \; f \; , \; r \; , \; n \right)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \ldots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented
- Action space: $\mathcal{A}(t, x, y) \subseteq \mathcal{A}$ is the set of admissible actions for (x, y) at t
  - Knapsack example: “select”; “do not select”
- State transition function: $f(t, x, y, a)$ state that one reaches for $a \in \mathcal{A}(t, x, y)$
  - Knapsack example: $f(t, x, y, \mathrm{select}) = \mathrm{x - y}$; $f(t, x, y, \mathrm{do\,not\,select}) = \mathrm{x}$
- Reward function: $r(t, x, y, a)$ reward for taking action a at time t when at (x, y)
  - Knapsack example: $f(t, x, y, \mathrm{select}) = 1$; $f(t, x, y, \mathrm{do\,not\,select}) = 0$
- Time horizon: $n < \infty$

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem
- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \le k \le n$$

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem
- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \le k \le n$$

- Expected total reward criterion, i.e. we are looking for $\pi_n^* \in \Pi(n)$ such that

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)].$$

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem
- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \leq k \leq n$$

- Expected total reward criterion, i.e. we are looking for $\pi_n^* \in \Pi(n)$ such that

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)].$$

- Dynamic Programming Recursion: for each $1 \leq t \leq n$ and for $x \in \mathcal{X}$,

$$v_t(x) = \mathbb{E}\left[\sup_{a \in \mathcal{A}(t,x,Y_t)} \{r(t, x, Y_t, a) + v_{t+1}(f(t, x, Y_t, a))\}\right],$$

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem
- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \leq k \leq n$$

- Expected total reward criterion, i.e. we are looking for $\pi_n^* \in \Pi(n)$ such that

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)].$$

- Dynamic Programming Recursion: for each $1 \leq t \leq n$ and for $x \in \mathcal{X}$,

$$v_t(x) = \mathbb{E}\left[\sup_{a \in \mathcal{A}(t,x,Y_t)} \{r(t, x, Y_t, a) + v_{t+1}(f(t, x, Y_t, a))\}\right],$$

  - $v_{n+1}(x) = 0$ for all $x \in \mathcal{X}$, and

MDPs: A Natural Context Where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem
- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \leq k \leq n$$

- Expected total reward criterion, i.e. we are looking for $\pi_n^* \in \Pi(n)$ such that

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)].$$

- Dynamic Programming Recursion: for each $1 \leq t \leq n$ and for $x \in \mathcal{X}$,

$$v_t(x) = \mathbb{E}\left[\sup_{a \in \mathcal{A}(t,x,Y_t)} \{ r(t, x, Y_t, a) + v_{t+1}\left(f(t, x, Y_t, a)\right) \} \right],$$

  - $v_{n+1}(x) = 0$ for all $x \in \mathcal{X}$, and
  - $v_1(\bar{x}) = \mathbb{E}[R_n(\pi_n^*)]$

Three Properties: Common and Easy to Check

Three Properties: Common and Easy to Check

Property (Bounded Rewards)

There is a constant $K < \infty$ *such that* $0 \leq r(t, x, y, a) \leq K$ *for all triples* (x, y, a) *and all times* $1 \leq t \leq n$.

Three Properties: Common and Easy to Check

Property (Bounded Rewards)

There is a constant $K < \infty$ *such that* $0 \leq r(t, x, y, a) \leq K$ *for all triples* (x, y, a) *and all times* $1 \leq t \leq n$.

Property (Existence of a Do-nothing Action)

For each time $1 \leq t \leq n$ *and pair* (x, y)*, the set of actions* $\mathcal{A}(t, x, y)$ *includes a do-nothing action* a^0 *such that*

$$r(t, x, y, a^0) = 0 \qquad \text{and} \qquad f(t, x, y, a^0) = x.$$

Three Properties: Common and Easy to Check

Property (Bounded Rewards)

There is a constant $K < \infty$ *such that* $0 \leq r(t, x, y, a) \leq K$ *for all triples* (x, y, a) *and all times* $1 \leq t \leq n$.

Property (Existence of a Do-nothing Action)

For each time $1 \leq t \leq n$ *and pair* (x, y)*, the set of actions* $\mathcal{A}(t, x, y)$ *includes a do-nothing action* a^0 *such that*

$$r(t, x, y, a^0) = 0 \qquad \text{and} \qquad f(t, x, y, a^0) = x.$$

Property (Optimal Action Monotonicity, or "Pay-to-Play")

For each time $1 \leq t \leq n$ *and state* $x \in \mathcal{X}$ *one has the inequality*

$$v_{t+1}(x^*) \leq v_{t+1}(x)$$

where $x^* = f(t, x, y, a^*)$ *and where* a^* *is an optimal action in* $\mathcal{A}(t, x, y)$.

A Variance Bounded by a Mean: Easy and Useful

A Variance Bounded by a Mean: Easy and Useful

Theorem (Arlotto, Gans, S., 2012)

Suppose that the Markov decision problem $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, f, r, n)$ *satisfies reward boundedness, existence of a do-nothing action and optimal action monotonicity. If* $\pi_n^* \in \Pi(n)$ *is a Markov deterministic policy such that*

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)],$$

then

$$\mathrm{Var}[R_n(\pi_n^*)] \le K\, \mathbb{E}[R_n(\pi_n^*)],$$

where K *is the uniform bound on the one-period reward function.*

A Variance Bounded by a Mean: Easy and Useful

Theorem (Arlotto, Gans, S., 2012)

Suppose that the Markov decision problem $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, f, r, n)$ *satisfies reward boundedness, existence of a do-nothing action and optimal action monotonicity. If* $\pi_n^* \in \Pi(n)$ *is a Markov deterministic policy such that*

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)],$$

then

$$\mathrm{Var}[R_n(\pi_n^*)] \leq K\, \mathbb{E}[R_n(\pi_n^*)],$$

where K *is the uniform bound on the one-period reward function.*

Corollary (Relative Stability)

If $\mathbb{E}[R_n(\pi_n^*)] \to \infty$ *as* $n \to \infty$*, then*

$$\frac{R_n(\pi_n^*)}{\mathbb{E}[R_n(\pi_n^*)]} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

Examples

Examples of MDPs that satisfy reward boundedness, existence of a do-nothing action and optimal action monotonicity:

- General dynamic and stochastic knapsack problems (Papastavrou, Rajagopalan and Kleywegt, 1996)
- Sequential investment problems (Derman et al., 1975; Prastacos, 1983)
- Capacity control problems in revenue management (Talluri and van Ryzin, 2004)
- Stochastic depletion problems with deterministic transitions (Chan and Farias, 2009)
- Sequential selection of monotone, unimodal and d-modal subsequences (Arlotto and S., 2011)
- More?

Bounding the Variance by the Mean: Sketch of the Proof

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \leq t \leq n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \leq t \leq n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

  is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$
- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \leq t \leq n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$

- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$
- For $d_t = M_t - M_{t-1}$,

$$\mathrm{Var}[M_n] = \mathrm{Var}\left[R_n(\pi_n^*)\right] = \mathbb{E}\left[\sum_{t=1}^{n} d_t^2\right]$$

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \leq t \leq n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$

- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$
- For $d_t = M_t - M_{t-1}$,

$$\mathrm{Var}[M_n] = \mathrm{Var}\left[R_n(\pi_n^*)\right] = \mathbb{E}\left[\sum_{t=1}^{n} d_t^2\right]$$

- An application of reward boundedness, existence of a do-nothing action, and optimal action monotonicity gives

$$\mathbb{E}[d_t^2 \mid \mathcal{F}_{t-1}] \leq K\, \mathbb{E}[r(t, X_t, Y_t, A_t^*) \mid \mathcal{F}_{t-1}]$$

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \leq t \leq n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$

- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$
- For $d_t = M_t - M_{t-1}$,

$$\mathrm{Var}[M_n] = \mathrm{Var}\left[R_n(\pi_n^*)\right] = \mathbb{E}\left[\sum_{t=1}^{n} d_t^2\right]$$

- An application of reward boundedness, existence of a do-nothing action, and optimal action monotonicity gives

$$\mathbb{E}[d_t^2 \,|\, \mathcal{F}_{t-1}] \leq K\, \mathbb{E}[r(t, X_t, Y_t, A_t^*) \,|\, \mathcal{F}_{t-1}]$$

- Taking total expectations and summing gives

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \leq K\, \mathbb{E}\left[R_n(\pi_n^*)\right]$$

Bounding the Variance by the Mean: Sketch of the Proof

- For $0 \le t \le n$, the process

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

  is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$

- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$

- For $d_t = M_t - M_{t-1}$,

$$\mathrm{Var}[M_n] = \mathrm{Var}\left[R_n(\pi_n^*)\right] = \mathbb{E}\left[\sum_{t=1}^{n} d_t^2\right]$$

- An application of reward boundedness, existence of a do-nothing action, and optimal action monotonicity gives

$$\mathbb{E}[d_t^2 \,|\, \mathcal{F}_{t-1}] \le K\,\mathbb{E}[r(t, X_t, Y_t, A_t^*) \,|\, \mathcal{F}_{t-1}]$$

- Taking total expectations and summing gives

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \le K\,\mathbb{E}\left[R_n(\pi_n^*)\right]$$

- Crucial here: $X_{t+1} = f(t, X_t, Y_t, A_t)$ is $\mathcal{F}_t$-measurable! □

Two Concrete Conjectures in the Simplest Context . . .

Two Concrete Conjectures in the Simplest Context ...

The simple knapsack problem is equivalent to the monotone subsequence problem:

- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent uniform on $[0, 1]$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max \left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq 1 \right\},$$

- π_n^*: optimal Markov deterministic policy such that $\mathbb{E}[R_n(\pi_n^*)] = \sup_\pi \mathbb{E}[R_n(\pi)]$

Two Concrete Conjectures in the Simplest Context . . .

The simple knapsack problem is equivalent to the monotone subsequence problem:

- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent uniform on $[0, 1]$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max \left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq 1 \right\},$$

- π_n^*: optimal Markov deterministic policy such that $\mathbb{E}\left[R_n(\pi_n^*)\right] = \sup_\pi \mathbb{E}\left[R_n(\pi)\right]$

Two Conjectures

Two Concrete Conjectures in the Simplest Context . . .

The simple knapsack problem is equivalent to the monotone subsequence problem:

- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent uniform on $[0, 1]$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq 1 \right\},$$

- π_n^*: optimal Markov deterministic policy such that $\mathbb{E}\left[R_n(\pi_n^*)\right] = \sup_\pi \mathbb{E}\left[R_n(\pi)\right]$

Two Conjectures

Two Concrete Conjectures in the Simplest Context ...

The simple knapsack problem is equivalent to the monotone subsequence problem:

- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent uniform on $[0, 1]$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{k : \sum_{i=1}^{k} Y_{\tau_i} \le 1\right\},$$

- π_n^*: optimal Markov deterministic policy such that $\mathbb{E}[R_n(\pi_n^*)] = \sup_\pi \mathbb{E}[R_n(\pi)]$

Two Conjectures

- From the MDP variance bound and Arlotto and S. (2011) [for the lower bound] we know

$$(1/3)\mathbb{E}[R_n(\pi_n^*)] - 2 \le \mathrm{Var}(R_n(\pi_n^*)) \le \mathbb{E}[R_n(\pi_n^*)] \qquad \text{for all } n \ge 1$$

and (prodded by a very insightful referee) we conjecture

$$\mathrm{Var}(R_n(\pi_n^*)) \sim (1/3)\mathbb{E}[R_n(\pi_n^*)] \quad \text{as } n \to \infty$$

Two Concrete Conjectures in the Simplest Context ...

The simple knapsack problem is equivalent to the monotone subsequence problem:

- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent uniform on $[0, 1]$
- Knapsack policy π: the number of items included is

$$R_n(\pi) = \max\left\{k : \sum_{i=1}^{k} Y_{\tau_i} \le 1\right\},$$

- π_n^*: optimal Markov deterministic policy such that $\mathbb{E}\left[R_n(\pi_n^*)\right] = \sup_\pi \mathbb{E}\left[R_n(\pi)\right]$

Two Conjectures

- From the MDP variance bound and Arlotto and S. (2011) [for the lower bound] we know

$$(1/3)\mathbb{E}\left[R_n(\pi_n^*)\right] - 2 \le \mathrm{Var}\left(R_n(\pi_n^*)\right) \le \mathbb{E}\left[R_n(\pi_n^*)\right] \qquad \text{for all } n \ge 1$$

  and (prodded by a very insightful referee) we conjecture

$$\mathrm{Var}\left(R_n(\pi_n^*)\right) \sim (1/3)\mathbb{E}\left[R_n(\pi_n^*)\right] \quad \text{as } n \to \infty$$

- CLT When centered and scaled $R_n(\pi_n^*)$ converges in distribution to $N(0, 1)$.

Not a Plan — Just Suggestions

"Take Aways: Hopefully Something New

Summary:

- The alphabet soup of an MDP can be off-putting, but the MDP structure is honestly rich and it is worth one's time to become familiar with it (if you've not done so already). You get all the "benefits" of abstraction.
- There is a natural martingale associated with any fixed-horizon MDP:

$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$

- This dynamic programming martingale is not as universal as the Doob martingale, but it still has substantial range. It is worth consideration in "any sequential problem"; this is good since in such problems the Doob martingale is often useless.
- Simple martingale arguments can be used to extract useful (but not quite precise) distributional information; the "trick" seems to rest in finding the features of the MDP that feedback into nice properties of the DP martingale.
- There are numerous open problems with a wide range of potential and of difficulty.

Muchas Gracias a Todos

References I

Arlotto, A. and Steele, J. M. (2011), 'Optimal sequential selection of a unimodal subsequence of a random sequence', *Combinatorics, Probability and Computing* **20**(06), 799–814.

Chan, C. W. and Farias, V. F. (2009), 'Stochastic depletion problems: effective myopic policies for a class of dynamic optimization problems', *Math. Oper. Res.* **34**(2), 333–350.

Coffman, Jr., E. G., Flatto, L. and Weber, R. R. (1987), 'Optimal selection of stochastic intervals under a sum constraint', *Adv. in Appl. Probab.* **19**(2), 454–473.

Derman, C., Lieberman, G. J. and Ross, S. M. (1975), 'A stochastic sequential allocation model', *Operations Res.* **23**(6), 1120–1130.

Papastavrou, J. D., Rajagopalan, S. and Kleywegt, A. J. (1996), 'The dynamic and stochastic knapsack problem with deadlines', *Management Science* **42**(12), 1706–1718.

Prastacos, G. P. (1983), 'Optimal sequential investment decisions under conditions of uncertainty', *Management Science* **29**(1), 118–134.

Talluri, K. T. and van Ryzin, G. J. (2004), *The theory and practice of revenue management*, International Series in Operations Research & Management Science, 68, Kluwer Academic Publishers, Boston, MA.