Non-Parametric Up-and-Down Experimentation
Revisited\textsuperscript{1}

Michael N. Katehakis

\textit{Rutgers Business School}

18TH ANNUAL APPLIED PROBABILITY DAY:
IN HONOR AND MEMORY OF CYRUS DERMAN

THE CENTER FOR APPLIED PROBABILITY AT COLUMBIA UNIVERSITY
December 2, 2011

\textsuperscript{1}Joint work with Cyrus Derman
The Basic Problem


- Let $Y(x)$ be a random variable such that:

$$
Y(x) = \begin{cases} 
1 & P(Y(x) = 1) = F(x) \\
0 & P(Y(x) = 0) = 1 - F(x)
\end{cases}
$$

where $F(x)$ is an unknown distribution function.

- **Objective**: given $\alpha$ estimate the $\alpha$-quantile of $F(x)$,

$$
x_\alpha = F^{-1}(\alpha)
$$

with observations distributed like $Y(x)$ where the choice of $x$ is under control.

- **Special Case**: Median: $x_{0.50} = L_{0.50}, \alpha = .50.$
Possible Cases

\[ F(x_\alpha) = \alpha \]

\[ x_\alpha = \inf \{ x \in \mathbb{N} : F(x) \geq \alpha \} \]

\[ x_\alpha = \sup \{ x \in \mathbb{N} : F(x) \leq \alpha \} \]
Possible Cases

\[ F(x_\alpha) = \alpha \]

\[ x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha\} \]

\[ x_\alpha = \sup\{x \in \mathbb{N} : F(x) \leq \alpha\} \]

Objective:

Start with \( x_0 \), observe \( Y(x_0) = 1, 0 \), \( \Pr(Y(x_0) = 1) = F(x_0) \).

choose \( X_1 \), observe \( Y(X_1) = 1, 0 \), \( \Pr(Y(X_1) = 1) = F(X_1) \).
Possible Cases

\[ F(x_\alpha) = \alpha \]

\[ x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha\} \]

\[ x_\alpha = \sup\{x \in \mathbb{N} : F(x) \leq \alpha\} \]

Objective:

Start with \( x_0 \), observe \( Y(x_0) = 1,0 \), \( \Pr(Y(x_0) = 1) = F(x_0) \).

Choose \( X_1 \), observe \( Y(X_1) = 1,0 \), \( \Pr(Y(X_1) = 1) = F(X_1) \).

\( \ldots, \ldots \)
Possible Cases

\[ F(x_\alpha) = \alpha \]

\[ x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha \} \]

\[ x_\alpha = \sup\{x \in \mathbb{N} : F(x) \leq \alpha \} \]

Objective:

Start with \( x_0 \), observe \( Y(x_0) = 1, 0, \Pr(Y(x_0) = 1) = F(x_0) \).

choose \( X_1 \), observe \( Y(X_1) = 1, 0, \Pr(Y(X_1) = 1) = F(X_1) \).

\( \ldots \), \( \ldots \)

choose \( X_n \), such that \( \hat{x}_\alpha(x_0, X_1, \ldots, X_n) \xrightarrow{a.s.} x_\alpha \).
Newsvendor Model

- When \( x \) items are ordered in a period we observe if there is:
  - no shortage \( Y(x) = 1 \) or shortage \( Y(x) = 0 \).
- The optimal order quantity \( x_\alpha \) is determined by a quantile requirement: \( x_\alpha = \inf \{ x \in \mathbb{N} : F(x) \geq \alpha \} = (p - c)/p \)
- \( F \) is the demand distribution.
Newsvendor Model

- When $x$ items are ordered in a period we observe if there is:
  - no shortage $Y(x) = 1$ or shortage $Y(x) = 0$.
- The optimal order quantity $x_\alpha$ is determined by a quantile requirement:
  $$x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha = (p - c)/p\}$$
- $F$ is the demand distribution.

Drug Testing

- Dosage is $x$
- Observe a success $Y(x) = 1$ or failure $Y(x) = 0$.
- The optimal dosage $x_\alpha$ is determined by a quantile requirement:
  $$x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha\}$$
- $F$ is the quantal response function.
Relevance

Newsvendor Model

- When $x$ items are ordered in a period we observe if there is: no shortage $Y(x) = 1$ or shortage $Y(x) = 0$.
- The optimal order quantity $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha = (p - c)/p \}$
- $F$ is the demand distribution.

Drug Testing

- Dosage is $x$
- Observe a success $Y(x) = 1$ or failure $Y(x) = 0$.
- The optimal dosage $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha \}$
- $F$ is the quantal response function.

Educational Testing

- Difficulty level of a test question is $x$
- Observe a correct $Y(x) = 1$ or wrong $Y(x) = 0$ answer.
- The student’s ability is the $x_\alpha$ for which $F(x_\alpha) = \alpha$
Newsvendor Model

- When $x$ items are ordered in a period we observe if there is:
  - no shortage $Y(x) = 1$ or shortage $Y(x) = 0$.
- The optimal order quantity $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha = (p - c)/p\}$
- $F$ is the demand distribution.

Drug Testing

- Dosage is $x$
- Observe a success $Y(x) = 1$ or failure $Y(x) = 0$.
- The optimal dosage $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha\}$
- $F$ is the quantal response function.

Educational Testing

- Difficulty level of a test question is $x$
- Observe a correct $Y(x) = 1$ or wrong $Y(x) = 0$ answer.
- The student’s ability is the $x_\alpha$ for which $F(x_\alpha) = \alpha$

Manufacturing
Relevance

Newsvendor Model

- When $x$ items are ordered in a period we observe if there is: no shortage $Y(x) = 1$ or shortage $Y(x) = 0$.
- The optimal order quantity $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha = (p - c)/p\}$
- $F$ is the demand distribution.

Drug Testing

- Dosage is $x$
- Observe a success $Y(x) = 1$ or failure $Y(x) = 0$.
- The optimal dosage $x_\alpha$ is determined by a quantile requirement: $x_\alpha = \inf\{x \in \mathbb{N} : F(x) \geq \alpha\}$
- $F$ is the quantal response function.

Educational Testing

- Difficulty level of a test question is $x$
- Observe a correct $Y(x) = 1$ or wrong $Y(x) = 0$ answer.
- The student’s ability is the $x_\alpha$ for which $F(x_\alpha) = \alpha$

Manufacturing

The Sensitivity of an Explosive
Derman’s Up & Down Method

• Grid or experimental range of \(x\) to a set of numbers of the form

\[
b + hn \quad (\infty < b < \infty, \ h > 0, \ n = 0, \pm 1, \ldots).
\]

For convenience one can assume \(b = 0, \ h = 1\).

• Procedure:
  Start with \(x_0\) (init. guess) \(y_0 = Y(x_0)\) is observed
  where \(P(Y(x_0) = 1) = F(x_0) = 1 - P(Y(x_0) = 0)\).
  Given \(x_0, y(x_0), \ldots, x_{n-1}, y(x_{n-1})\) for \(n \geq 0\), define:

\[
x_{n+1} = \begin{cases} 
  x_n + 1 & \text{if } y(x_n) = 0 \text{ with probability: } 1 \\
  x_n - 1 & \text{if } y(x_n) = 1 \text{ with probability: } 1 - \frac{1}{2\alpha} \\
  x_n & \text{if } y(x_n) = 1 \text{ with probability: } \frac{1}{2\alpha}
\end{cases}
\]

where w.l.o.g. \(\alpha > 1/2\)

• The estimate \(\hat{x}_{\alpha,n}\) of \(x_\alpha\) based on \(n\) observations is

\[
\hat{x}_{\alpha,n} = \begin{cases} 
  \text{the most frequent value of } x'\text{s,} & \text{if unique,} \\
  \text{the arithmetic average of the most frequent levels} & \text{otherwise.}
\end{cases}
\]
Derman’s Main Result

\[{X_n, \}_{n \geq 1} \text{ process: } p_{i,i+1} = 1 - F(i)/2\alpha\]
Derman’s Main Result

\[ \pi(x) = \lim_{n} Pr(X_n = x | x_0) \]

\[ \pi(0) \leq \pi(1) \leq \ldots \leq \pi(x_\alpha - 1) < \pi(x_\alpha) \geq \pi(x_\alpha + 1) \geq \ldots \]

\[ P_{x,.-1:r,.-1} \]

\[ \cdots \]

Derman’s Main Result

\[ \{X_n, \}_{n \geq 1} \text{ process: } p_{i,i+1} = 1 - F(i)/2\alpha \]

Condition A: If \( F(x) \) is strictly increasing for \( x \in [x_\alpha - 1, x_\alpha] \) then:

\[ \Pr(max(|\lim_{n} \hat{x}_{\alpha,n} - x_\alpha|, |\lim_{n} \hat{x}_{\alpha,n} - x_\alpha|) < 1) = 1 \]
Derman’s Main Result

\[ \{ X_n, \}_{n \geq 1} \] process: \( p_{i,i+1} = 1 - F(i)/2\alpha \)

**Condition A:** If \( F(x) \) is strictly increasing for \( x \in [x_\alpha - 1, x_\alpha] \) then:

\[ \Pr(\max(|\lim_{n} \hat{x}_{\alpha,n} - x_\alpha|, |\lim_{n} \hat{x}_{\alpha,n} - x_\alpha|) < 1) = 1 \]

**Main Tool:** \( \pi(x) = \lim_{n} \Pr(X_n = x | x_o) \)

\( \pi(0) \leq \pi(1) \leq \ldots \leq \pi(x_\alpha - 1) < \pi(x_\alpha) \geq \pi(x_\alpha + 1) \geq \pi(x_\alpha + 2) \geq \ldots \)
Derman’s U-D Revisited

\[ \{ X_n, \} \text{ process: } p_{i,i+1} = 1 - F(i)/2\alpha \]

- What is an efficient **Stopping Criterion**?
- What is the **Error Probability**?
- **What if Condition A does not hold**?
Observation 1 if $x^\alpha$ is on the grid:

$p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x^\alpha - 1}x^\alpha \geq p_{x^\alpha}x^\alpha + 1 = 1/2 \geq p_{x^\alpha}x^\alpha + 1 \geq \ldots$

Observation 2 if $x^\alpha$ is not on the grid:

$p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x^\alpha - 1}x^\alpha > 1/2 \geq p_{x^\alpha}x^\alpha + 1 \geq p_{x^\alpha}x^\alpha + 2 \geq \ldots$

Compare with:

$\pi(0) \leq \pi(1) \leq \ldots \leq \pi(x^\alpha - 1) \leq \pi(x^\alpha) \geq \pi(x^\alpha + 1) \geq \pi(x^\alpha + 2) \geq \ldots$

{\{X_n, \}_{n \geq 1} \text{ process: } p_{i,i+1} = 1 - F(i)/2\alpha}
Observation 1 if $x_\alpha$ is on the grid

$$p_{00} = 1 \geq p_{12} \geq \cdots \geq p_{x_\alpha - 1 x_\alpha} \geq p_{x_\alpha x_\alpha + 1} = \frac{1}{2} \geq p_{x_\alpha + 1 x_\alpha + 2} \geq \cdots$$
Derman’s U-D Revisited - Answers

\{X_n, \}_{n \geq 1} \text{ process: } p_{i,i+1} = 1 - F(i)/2\alpha

Observation 1 if \(x_\alpha\) is on the grid

\[ p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x_\alpha-1x_\alpha} \geq p_{x_\alpha x_\alpha+1} = \frac{1}{2} \geq p_{x_\alpha+1x_\alpha+2} \geq \ldots \]

Observation 2 if \(x_\alpha\) is not on the grid:

\[ p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x_\alpha-1x_\alpha} > \frac{1}{2} \geq p_{x_\alpha x_\alpha+1} \geq p_{x_\alpha+1x_\alpha+2} \geq \ldots \]
Derman’s U-D Revisited - Answers

\{X_n, \}_{n \geq 1} \text{ process: } p_{i, i+1} = 1 - F(i)/2\alpha

Observation 1 if \( x_\alpha \) is on the grid

\[ p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x_\alpha - 1 x_\alpha} \geq p_{x_\alpha x_\alpha + 1} = 1/2 \geq p_{x_\alpha + 1 x_\alpha + 2} \geq \ldots \]

Observation 2 if \( x_\alpha \) is not on the grid:

\[ p_{00} = 1 \geq p_{12} \geq \ldots \geq p_{x_\alpha - 1 x_\alpha} > 1/2 \geq p_{x_\alpha x_\alpha + 1} \geq p_{x_\alpha + 1 x_\alpha + 2} \geq \ldots \]

Compare with :

\[ \pi(0) \leq \pi(1) \leq \ldots \leq \pi(x_\alpha - 1) \leq \pi(x_\alpha) \geq \pi(x_\alpha + 1) \geq \pi(x_\alpha + 2) \geq \ldots \]
\( \alpha = .80, \ x_\alpha = 24 \) on the grid: \( F(24) = .8 \)

\[ p_{23,24} > \frac{1}{2} = p_{24,25} > p_{25,26} < \frac{1}{2} \]
α = .80, x_α = 24 not on the grid: \( F(23) < 0.8 < F(24) \)

\[ p_{23,24} > 1/2 > p_{24,25} > p_{25,26} > \ldots \]
For simplicity consider the case $x_\alpha$ is on the grid.

- We can **Stop** the procedure:
  - $\tau_1^\alpha = \inf\{k : \hat{p}_{k,k+1}(x_0, \ldots, X_k) \in (1/2 - \epsilon, 1/2 + \epsilon)\}$
  - $\tau_2^\alpha = \inf\{k : \hat{p}_{k,k+1}(x_0, \ldots, X_k) \in (1/2 - \epsilon, 1/2 + \epsilon) \& V_k = \max_{k'} \{V_{k'}\} \}$
  - $\hat{x}_\alpha = X_{\tau_1^\alpha}$
  - Have $\Pr(\tau_\alpha^i > u) = c_F^i \left(e^{-u} + \epsilon_F^i(n, u)\right)$
    where $\epsilon_F^i(n, u) \to 0 \ (n \to \infty)$
  - We can modify the procedure by taking a second sample at $\tau_\alpha$ (Two Stage)
  - Working on using techniques of Adaptive MDPs to obtain:
    $$R^\pi_N \geq R^{DK}_N = M_{DK}(P)\log N + o(\log N)$$
For simplicity consider the case $x_\alpha$ is on the grid.

- We can **Stop** the procedure:
  - $\tau_1^\alpha = \inf\{k : \hat{p}_{k,k+1}(x_0, \ldots, X_k) \in (1/2 - \epsilon, 1/2 + \epsilon)\}$
  - $\tau_2^\alpha = \inf\{k : \hat{p}_{k,k+1}(x_0, \ldots, X_k) \in (1/2 - \epsilon, 1/2 + \epsilon) \& V_k = \max_{k'} \{V_{k'}\} \}$
  - $\hat{x}_\alpha = X_{\tau_1^\alpha}$
- Have $\Pr(\tau_1^\alpha > u) = c_F^i(e^{-u} + \epsilon_F^i(n, u))$
  where $\epsilon_F^i(n, u) \to 0$ ($n \to \infty$)
- We can modify the procedure by taking a second sample at $\tau_\alpha$ (Two Stage)
- Working on using techniques of Adaptive MDPs to obtain:
  \[
  R_N^\pi \geq R_N^{\pi_{DK}} = M_{DK}(P)\log N + o(\log N)
  \]

Similar results hold in the case $x_\alpha$ is not on the grid or Condition A does not hold.
Background

Non Parametric “Up & Down” Methods

- Derman C. (1957). Non-Parametric Up-and-down Experimentation
Background

Non Parametric “Up & Down” Methods

- Derman C. (1957). Non-Parametric Up-and-down Experimentation
- Ivanova A(2006). Dose-Finding in Oncology-Nonparametric Methods,
Background: “Stochastic Approximation” and Related Methods


The Robbins and Monro SA scheme can be used for estimating any quantile and it imposes no parametric assumptions on $F(x)$.

- The method does assume, however, that the range of possible experimental values of $x$ is the real line.
- It has slow convergence rate.
Background: Optimal Adaptive Policies for MABs


Note: $R_{\pi N} = M_{LR}(P) \log N + o(\log N)$

Even better: $\lim_{N \to \infty} \frac{R_{\pi 0 N}}{R_{\pi N}} \leq 1$

Lai T.L. and H. E. Robbins (1985) "Asymptotically Efficient Adaptive Allocation Rules", $\pi N = \max\{\mu_1, \ldots, \mu_n\}$
Background: Optimal Adaptive Policies for MABs


\[
R_N^\pi = \max\{\mu_1, \ldots, \mu_n\} N - V_N^\pi
\]

\[
R_N^\pi \geq R_N^{LR} = M_{LR}(P)\log N + o(\log N)
\]

- Note:

\[
R_N^{0\pi} = M_{LR}(P)\log N + o(\log N) = o(N^a) \quad \forall a > 0
\]

- Even better:

\[
\lim_{N} R_N^{0\pi} / R_N^{\pi} \leq 1
\]


### Equations:

1. \[ R_N^\pi = \max\{\mu_1, \ldots, \mu_n\} N - V_N^\pi \]
2. \[ R_N^\pi \geq R_N^{LR} = M_{LR}(P)\log N + o(\log N) \]
3. \[ R_N^{0\pi} = M_{LR}(P)\log N + o(\log N) = o(N^a) \quad \forall a > 0 \]
4. \[ \lim_{N} R_N^{0\pi} / R_N^{\pi} \leq 1 \]
Background: Optimal Adaptive Policies for MDPs


\[ R_N^\pi = V_N - V_N^{\pi} \]
\[ R_N^\pi \geq R_N^{\pi BK} = M_{BK}(P)\log N + o(\log N) \]
Background: Optimal Adaptive Policies for MDPs


\[ R_N^\pi = V_N - V_N^\pi \]

\[ R_N^\pi \geq R_N^{BK} = M_{BK}(P)\log N + o(\log N) \]

Near-optimal Regret Bounds for Reinforcement Learning


\[ R_N^\pi \geq R_N^{AO} = M_{AO}(P)\log N + o(\log N) \]


\[ R_N^\pi \geq R_N^{TB} = M_{TB}(P)\log N + o(\log N) \]

Background: Optimal Adaptive Policies for MDPs


\[ R^\pi_N = V_N - V^\pi_N \]

\[ R^\pi_N \geq R^\pi_{BK} = M_{BK}(P)\log N + o(\log N) \]

Near-optimal Regret Bounds for Reinforcement Learning


\[ R^\pi_N \geq R^\pi_{AO} = M_{AO}(P)\log N + o(\log N) \]


\[ R^\pi_N \geq R^\pi_{TB} = M_{TB}(P)\log N + o(\log N) \]


\[ M_{AO}(P) > M_{TB}(P) > M_{BK}(P) \forall P \]
Background: Optimal Adaptive Policies for MDPs


\[ R_N^\pi = V_N - V_N^\pi \]

\[ R_N^\pi \geq R_N^{BK} = M_{BK}(P)\log N + o(\log N) \]

Near-optimal Regret Bounds for Reinforcement Learning


\[ R_N^\pi \geq R_N^{AO} = M_{AO}(P)\log N + o(\log N) \]


\[ R_N^\pi \geq R_N^{TB} = M_{TB}(P)\log N + o(\log N) \]


\[ M_{AO}(P) > M_{TB}(P) > M_{BK}(P) \forall P \]
Background: Optimal Adaptive Policies for MDPs


\[
R_N^\pi = V_N - V_N^\pi
\]

\[
R_N^\pi \geq R_N^{BK} = M_{BK}(P)\log N + o(\log N)
\]

Near-optimal Regret Bounds for Reinforcement Learning


\[
R_N^\pi \geq R_N^{AO} = M_{AO}(P)\log N + o(\log N)
\]


\[
R_N^\pi \geq R_N^{TB} = M_{TB}(P)\log N + o(\log N)
\]


\[
M_{AO}(P) > M_{TB}(P) > M_{BK}(P) \forall P
\]
The Robbins Monro Method for $x_\alpha$

Compute $x_\alpha$ such that

$$M(x_\alpha) = \alpha$$

- **Deterministic case** $M(x)$ is **known** solution by:

  $$x_n = x_{n-1} + a_n[\alpha - M(x_{n-1})],$$

- **Stochastic case** $M(x)$ is **unknown**

  $$E(Y(x)) = M(x) = \alpha. \ \forall \ x$$

  Solution by:

  $$x_n = x_{n-1} + a_n[\alpha - y_{n-1}],$$

The estimate $\hat{x}_{\alpha,n}$ of $x_\alpha$ based on $n$ observations is $\hat{x}_{\alpha,n} = x_n$,

*Under regularity conditions*\(^2\)

\(^2\)E.g., $M$ is non-decreasing, there exists a solution $M(x_\alpha) = \alpha$, $\exists \frac{M(x)}{dx} > 0$ at $x_\alpha$, and

$$\sum_{n=0}^{\infty} a_n = \infty, \ \sum_{n=0}^{\infty} a_n^2 < \infty$$
The Dixon and Mood Method for $LD_{50}(x_{0.50})$

- Grid or experimental range$^3$ of $x$ to a set of numbers of the form

$$b + hn \ (−∞ < b < ∞, \ h > 0, \ n = 0, ±1, \ldots)$$

For convenience one can assume $b = 0$, $h = 1$.

- Data: Treatments are administered sequentially at dosage: $X_i$, as follows:
  Start with $x_0$ (arbitrary guess) $y_0 = Y(x_0)$ is observed where $P(Y(x_0) = 1) = F(x_0) = 1 - P(Y(x_0) = 0)$.
  Given $x_0, y(x_0), \ldots, x_k, y(x_{n−1})$ for $n ≥ 0$, define recursively

$$x_n = \begin{cases} 
  x_{n−1} + 1 & \text{if } y(x_k) = 0, \\
  x_{n−1} - 1 & \text{if } y(x_k) = 1.
\end{cases}$$

- The estimate $\hat{x}_α$ of $x_α$ based on $n$ observations is

$$\hat{x}_{0.50} = \frac{1}{n + 1} \sum_{j=1}^{n+1} x_j,$$

$^3$Natural limitations such as when $x$ is obtained by a counting procedure - limitations on the precision of measuring instruments.