# A Minimally Conservative Indifference Zone Policy

Peter I. Frazier

Operations Research & Information Engineering, Cornell University

Sunday November 7, 2010
INFORMS Annual Meeting
Austin, Texas

# Ranking and Selection

- We have $k$ "alternatives" or "systems" that can be simulated.
  - e.g., different methods for operating a supply chain.
- Each time we simulate alternative $x$, we observe

$$y \sim \mathrm{Normal}(\theta_x, \sigma_x^2) \qquad \text{(independent across } x\text{)}$$

  where $\theta_x$ is unknown. (This talk assumes known constant $\sigma_x^2 = \sigma^2$).
- Goal: Use simulation efficiently to find $\arg\max_x \theta_x$.

# Ranking and Selection Policies

- A policy is a rule for deciding
  - Which alternatives to sample at each point in time.
  - When to stop sampling.
  - Which alternative to select as the best when we stop sampling. [Policies usually select the one with the largest sample mean]
- These decisions can be made adaptively.
- Given a system configuration $\theta = (\theta_1, \ldots, \theta_k)$, and a policy $\pi$,

$$\mathrm{PCS}(\pi, \theta) \qquad \text{(Probability of Correct Selection)}$$

is the probability that $\pi$ selects an alternative in $\arg\max_x \theta_x$.

# Preference and Indifference Zones

- The **preference zone** is a collection of system configurations where the best is better than the second best by at least $\delta$ :

$$\mathrm{PZ}(\delta) = \left\{ \theta \in \mathbb{R}^k : \theta_{[1]} - \theta_{[2]} \geq \delta \right\},$$

  where $\theta_{[1]} \geq \theta_{[2]} \geq \ldots \geq \theta_{[k]}$ are the ordered components of $\theta$, and $\delta > 0$ is a fixed parameter.
- The **indifference zone** are those system configurations outside the preference zone.
- We say a policy $\pi$ has an **IZ guarantee** with parameters $\delta$ and $P^*$ if

$$\mathrm{PCS}(\pi, \theta) \geq P^* \quad \text{for all } \theta \in \mathrm{PZ}(\delta).$$

# Indifference Zone Formulation

Goal: Find a policy that satisfies the IZ guarantee, **while taking as few samples as possible**. Lots of previous work constructing policies that satisfy the IZ guarantee.

- Fixed sample size policies: [Bechhofer, 1954]
- Two-stage policies: [Dudewicz and Dalal, 1975, Rinott, 1978]
- Fully sequential policies [Paulson, 1964, Bechhofer and Goldsman, 1987, Hartmann, 1988, Hartmann, 1991, Paulson, 1994, Kim and Nelson, 2001, Nelson et al., 2001, Hong, 2006]

# Controllability

Many IZ policies are "difficult to control."

- Let $\pi$ be a policy that is known to satisfy the IZ guarantee with parameters $\delta$ and $P^*$.
- Let $\mathrm{PCS}(\pi,*) = \inf_{\theta \in \mathrm{PZ}(\delta)} \mathrm{PCS}(\pi,\theta)$ be the actual worst PCS in the preference zone.
- $\mathrm{PCS}(\pi,*) - P^*$ is the additional unrequested PCS. It is nonnegative.
- Existing fully sequential policies are **difficult to control** when $k$ is large: $\mathrm{PCS}(\pi,*) - P^* > 0$ is large enough to induce **significant and undesired additional sampling effort**.
- Goal: $\mathrm{PCS}(\pi,*) = P^*$.

# Maximum Controllability

- We now construct a fully sequential continuous-time IZ policy called **MCIZ-C** that is maximally controllable, i.e.,

$$\mathrm{PCS}(\text{MCIZ-C}, *) = P^*.$$

- A discrete-time version of this policy, called **MCIZ-D**, also satisfies the IZ guarantee. In addition,

$$\mathrm{PCS}(\text{MCIZ-D}, *) \approx P^*.$$

  The bound is not exact because of "overshoot" when stopping.
- MCIZ-C stands for "maximimally controllable IZ - continuous-time", and the D in "MCIZ-D" stands for "discrete-time"

# Construction of the MCIZ Policy

- MCIZ samples every alternative at each time $t$ (equal allocation).
  - It is a procedure "without elimination."
- Upon stopping, MCIZ selects the alternative with the largest sample mean as the best.
- The innovation in MCIZ is in the curved stopping boundary it uses to decide when to stop.

# Construction of the MCIZ Policy

- The MCIZ policy is derived using Bayesian ideas (!)
- Let $Q$ be a prior probability measure on $\mathrm{PZ}(\delta)$ under which

$$
\begin{aligned}
X_* &\sim \mathrm{Uniform}(1, \ldots, k) \\
a &\sim \mathrm{Uniform}(\mathbb{R}) \\
\theta_x &= \begin{cases} a + \delta & \text{if } x = X_* \\ a & \text{if } x \neq X_* \end{cases}
\end{aligned}
$$

- $Q$ is concentrated on the least-favorable configurations.

# Bayesian Posterior

- Let $Y_t = (Y_{t1}, \ldots, Y_{tk})$, where $Y_{tx}$ is the sum of all observations from alternative $x$ by time $t$.
  - In discrete time, $Y_{tx}$ is a random walk with independent $\mathrm{Normal}(\theta_x, \sigma^2)$ increments.
  - In continuous time, $Y_{tx}$ is a Brownian motion with drift $\theta_x$ and volatility $\sigma^2$.
- Given data $Y_t$, the Bayesian posterior distribution on the best alternative $X_*$ is

$$q_{tx} = Q\{X_* = x \mid Y_t\} = \frac{\exp\left(\frac{\delta}{\sigma} Y_{tx}\right)}{\sum_{x'=1}^{k} \exp\left(\frac{\delta}{\sigma} Y_{tx'}\right)}$$

# MCIZ Policy

- $q_{tx}$ is the Bayesian posterior probability that alternative $x$ is best.
- If we stop at time $t$, our Bayesian probability of correct selection is $\max_x q_{tx}$. [Different than the (frequentist) PCS already discussed].
- The MCIZ policy stops sampling when this Bayesian probability of correct selection exceeds our target $P^*$. This stopping time, call it $\tau$, can be written:

$$\tau = \inf\left\{t \in \mathbb{R}_+ : \max_x q_{tx} \geq P^*\right\} \quad \text{in continuous time (MCIZ-C)},$$

$$\tau = \inf\left\{t \in \mathbb{Z}_+ : \max_x q_{tx} \geq P^*\right\} \quad \text{in discrete time (MCIZ-D)}.$$

# Main Result

## Theorem

*The MCIZ policy, in both continuous and discrete time, satisfies the IZ guarantee, i.e.,*

$$\mathrm{PCS}(\text{MCIZ-C}, \theta) \geq P^* \quad \forall \theta \in \mathrm{PZ}(\delta)$$
$$\mathrm{PCS}(\text{MCIZ-D}, \theta) \geq P^* \quad \forall \theta \in \mathrm{PZ}(\delta).$$

## Theorem

*The MCIZ policy in continuous time is maximally controllable, i.e.,*

$$\inf_{\theta \in \mathrm{PZ}(\delta)} \mathrm{PCS}(\text{MCIZ-C}, \theta) = P^*$$

Recall $\inf_{\theta \in \mathrm{PZ}(\delta)} \mathrm{PCS}(\mathrm{MCIZ} - \mathrm{C}, \theta) = \mathrm{PCS}(\mathrm{MCIZ} - \mathrm{C}, *)$.

# Proof Sketch

For $\theta \in \mathbb{R}^d$, let $Q_\theta$ be a prior that is uniform on the permutations and translations of $\theta$. In particular, $Q = Q_{[\delta,0,\ldots,0]}$.

## Lemma (Symmetry)

$\mathrm{PCS}(\pi,\theta) = Q^\pi_\theta\{\mathrm{CS}\}$ *for* $\pi =$ *MCIZ-C or* $\pi =$ *MCIZ-D.*

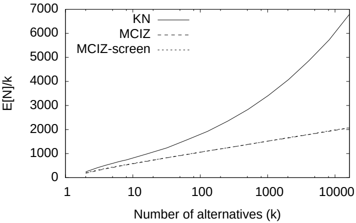
## Lemma (Monotonicity)

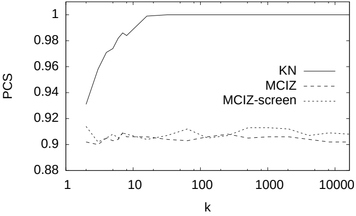
*Let* $q_{tx}(\theta)$ *be the posterior probability under* $Q_\theta$ *that alternative* $x$ *is best. Then, for any* $Y_t$ *and any* $\theta \in \mathrm{PZ}(\delta)$, $\max_x q_{tx}(\theta) \geq \max_x q_{tx}$.

## Lemma (Stopping)

*For* $\pi =$ *MCIZ-D,* $Q^\pi_\theta\{\mathrm{CS}\} \geq P^*$.
*For* $\pi =$ *MCIZ-C,* $Q^\pi_\theta\{\mathrm{CS}\} = P^*$.

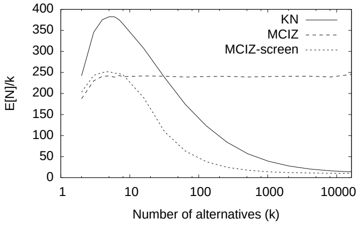
# Numerical Comparisons: Slippage Configuration

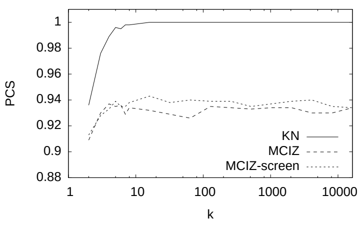



- The slippage configuration is $\theta = [0, \ldots, 0, \delta]$.
- KN is a fully sequential policy from [Kim & Nelson 2001]
- Evaluation is in discrete time, with $P^* = 0.9$, $\sigma = 10$, $\delta = 1$, estimated with 1000 independent replications.

# Numerical Comparisons: Monotone Decreasing Means




- The monotone decreasing means configuration is $\theta = [\delta, 2\delta, \ldots, k\delta]$.
- MCIZ-screen is a version of MCIZ that spends some error probability in a first stage that screens out clearly poor alternatives.
- Evaluation is in discrete time, with $P^* = 0.9$, $\sigma = 10$, $\delta = 1$, estimated with 1000 independent replications.

# Conclusion

- The MCIZ-C policy is a fully sequential policy IZ policy that is **maximally controllable**.
- To my knowledge, this is the **first fully sequential IZ policy with this property** for $k > 2$.
- This method of Bayesian analysis with least-favorable priors is a general theoretical tool.
- Ongoing work:
  - Generalize beyond known common variance.
  - Generalize beyond equal allocation, or allow dropping alternatives.

# Thank You; Any Questions?

- If you are interested in these topics, please consider submitting a paper to an upcoming **special issue of IIE Transactions** devoted to simulation optimization and its applications.
- Due Date for Submission: June 2011
- Special Issue Editors: Loo Hay Lee; Ek Peng Chew; Samuel Qing-Shan Jia; Peter Frazier; Chun-Hung Chen

# References


Bechhofer, R. (1954).
A single-sample multiple decision procedure for ranking means of normal populations with known variances.
*The Annals of Mathematical Statistics*, 25(1):16–39.

Bechhofer, R. and Goldsman, D. (1987).
Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the normal population which has the largest mean.
*Communications in Statistics-Simulation and Computation*, 16(4):1067–1092.

Dudewicz, E. and Dalal, S. (1975).
Allocation of observations in ranking and selection with unequal variances.
*Sankhyā: The Indian Journal of Statistics, Series B*, 37(1):28–78.

Hartmann (1991).
An improvement on paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance.
*Sequential Analysis*, 10(1-2):1–16.

Hartmann, M. (1988).
An improvement on paulson s sequential ranking procedure.
*Sequential Analysis*, 7(4):363–372.

Hong, J. (2006).
Fully sequential indifference-zone selection procedures with variance-dependent sampling.
*Naval Research Logistics (NRL)*, 53(5):464–476.

Kim, S.-H. and Nelson, B. L. (2001).
A fully sequential procedure for indifference-zone selection in simulation.
*ACM Trans. Model. Comput. Simul.*, 11(3):251–273.

Nelson, B., Swann, J., Goldsman, D., and Song, W. (2001).
Simple procedures for selecting the best simulated system when the number of alternatives is large.
*Operations Research*, 49(6):950–963.

# Continuation Region

- Images show continuation region for $k = 3$,
  in linear coordinates (left) and exponential coordinates (right).
- MCIZ stops when $Y_t$ exits the continuation region.