Bandit Problems and Adaptive Clinical Trials

Xikui Wang, PhD

Department of Statistics

University of Manitoba

Winnipeg, Manitoba

Canada R3T 2N2

Workshop on Adaptive Designs

The Fields Institute, Toronto, Canada

September 27, 2003

Acknowledgement: supported by NSERC

Objectives

1. To introduce the unified decision theoretic approach for various types of trials from both ethical and mathematical points of view, and to motivate the use of adaptive designs.

2. To make the connection between adaptive designs and bandit processes.

3. To discuss some recent results of bandit processes with delayed responses.

Some terms

- CCTs: controlled clinical trials
- RCTs: randomized clinical trials
- ACTs: adaptive clinical trials
- SCTs: sequential clinical trials

1. Controlled Clinical Trials (CCTs)

1). The setting

- treatments: several (normally two) alternative medical interventions for a common disease, with unknown effectiveness

- horizon: an unknown number N of patients with the common disease, to be treated by one and only one intervention

- responses: immediate or delayed

- decisions: treatment allocation and trial termination

$\boxed{\text{1. Controlled Clinical Trials (CCTs)}}$

2). Philosophical viewpoints

(Edwards et al. 1998)

- utilitarian

"... one's ultimate duty is to maximise utility by producing happiness of the greatest number of people - all other duties being derived from this."

- Kantian

"... one should always treat people with respect - **never** treating them merely as the means to other people's ends."

- the competing interests

trial participants and the society

1. Controlled Clinical Trials (CCTs)

3). Ethical issues (Clayton 1982)

- collective ethics / common good

"It is the duty of the doctor to acquire new knowledge so that, by such advance, future patients might benefit, ..."

- individual ethics / personal care

"It is the duty of the doctor to apply existing knowledge for the best possible treatment of each individual patient."

- the ethical dilemma: competing duties

information gathering versus immediate payoff

1. Controlled Clinical Trials (CCTs)

4). Statistical issues (Simon 1991)

- statistical design and analysis
- trial termination: sample size
- treatment allocation
- statistical interim analysis
- control of confounding covariates

1. Controlled Clinical Trials (CCTs)

5). Practical issues

- recruitment of patients

- "truly" informed consents from patients

- clinicians' collaboration

- multi-centre trials

- data monitoring committee

- cost and management

	God	Devil
patient	physician	Randomization
physician	$p < 0.05$	$p > 0.05$
statistician	statistician	$n = 1$

2. Randomized Clinical Trials (RCTs)

1). Current state of the art

- the gold standard

- religion "trialism" (Rimm & Bortin 1978)

- a "hallowed status" (Berry 1989)

- "[s]ome biostatisticians and clinicians refuse to believe that a treatment has an effect unless is has been shown in a 'properly conducted' randomized clinical trial." (Berry 1989)

- "... it remains an ideal that all new healthcare interventions should be evaluated through randomized controlled trials" (Sibbald and Roland 1998)

2. Randomized Clinical Trials (RCTs)

2). Problem 1: unethical randomization

Example 1 - antiviral zidovudine treatment (AZT) trial: reducing the risk of maternal-to-infant HIV transmission

(Connor et al 1994, Rosenberger 1996)

Treatment	total	HIV+
AZT	238	20
Placebo	238	60

- a simulation study (Yao and Wei, 1996)

$\boxed{\text{2. Randomized Clinical Trials (RCTs)}}$

Problem 1: unethical randomization

Example 2 - extracorporeal membrane oxygenation (ECMO) trials: a cardiopulmonary bypass treatment for severe but potentially reversible persistent pulmonary hypertension of the newborn (PPHN)

(Bartlett et al 1985): ACT, RPW

(O'Bourke et al 1989): 2-stage SCT

(Gross et al 1994): RCT

(UK Collaborative 1996): RCT

- post study analysis of UK ECMO trial (Snowdon et al 1997)

2. Randomized Clinical Trials (RCTs)

Problem 2: infeasible randomization

- clinicians declined to recruit patients for randomized allocation (Fairhurst and Dowrick 1996)

- strong patient preferences (Brewin and Bradley 1989; Emanuel and Patterson 1998)

3. Adaptive Clinical Trials (ACTs)

1). Moral requirement of ACTs

- the dual role and responsibility of the researcher/clinician

- the dual role and contribution of the subject/patient

- *Declaration of Helsinki*:

"the interests of science and society should *never* take precedence over consideration related to the well-being of the subject" (*World Medical Assembly*, 1996)

- informed consent infeasible in desperate medical situations

3. Adaptive Clinical Trials (ACTs)

2). Ethical justification for ACTs

- the principle of interchangeability (PoI): any two patients are ethically interchangeable (that is, at the point of enrollment in a clinical trial, the intent is to provide the best treatment available to each patient given current information) (Pullman and Wang 2001)

- RCTs (collective ethics): fail the PoI

- myopic allocation (individual ethics): fails the PoI

- ACTs: satisfy the PoI

$\boxed{\text{4. RCTs and ACTs: unification}}$

1). A decision theoretic model:

- a common ground: allocate the better treatment to "*more*" patients

- strategy $\pi = (\pi_1, \cdots, \pi_n, \cdots)$: trial termination and treatment allocation

- the worth of the strategy π:

$$E_\pi(Z_1 + Z_2 + \cdots + Z_N)$$

- Z_i: response from the i^{th} patient

- objective: maximize $E_\pi(Z_1 + \cdots + Z_N)$

4. RCTs and ACTs: unification

2). The bandit processes formulation

(Berry and Fristedt 1985)

- unknown N follows geometric$(1-\alpha)$
- objective: maximize $E_\pi(\Sigma_{n=1}^{\infty}\alpha^{n-1}Z_n)$
- treatments: i.i.d. $X_{i,n} \sim F_i$
- assumption: (F_1, F_2) is unknown
- treatment: $\pi_n \in \{1,2\}$ for n^{th} patient
- response: $Z_n = X_{\pi_n,n}$ for n^{th} patient
- approach: Bayesian (Markov decision processes, dynamic programming)
- essential feature: information gathering and immediate payoff
- satisfies PoI

$\boxed{\text{4. RCTs and ACTs: unification}}$

3). Just a mathematical generalization!

- domains of strategies:

$$\Pi_{RCTs} \subset \Pi_{SCTs} \subset \Pi_{ACTs}$$

design	trial termination	treatment allocation
RCTs	No	No
SCTs	Yes	No
ACTs	Yes	Yes

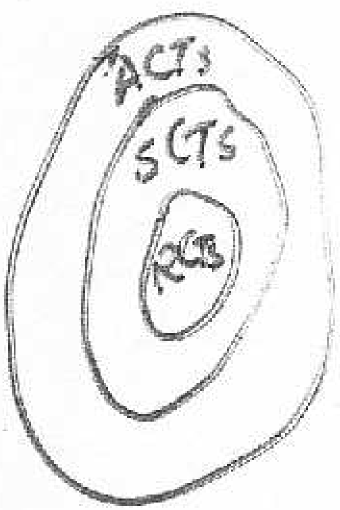
- ACTs: one or both of clinical decisions depend on accumulating information, and include SCTs as special cases

4. RCTs and ACTs: unification

4). A comparison - minimax approach (Wang and Pullman 2001)

- responses: immediate and dichotomous
- probabilities of successes: P_A and P_B
- regret of successes lost: $R_\pi(P_A, P_B)$

$$= N \max\{P_A, P_B\} - E_\pi(Z_1 + \cdots + Z_N)$$



$$\inf_{\pi \in \Pi_{ACTs}} R_\pi(P_A, P_B)$$

$$< \inf_{\pi \in \Pi_{SCTs}} R_\pi(P_A, P_B)$$

$$< \inf_{\pi \in \Pi_{RCTs}} R_\pi(P_A, P_B)$$

5. Bandit Processes

1). Major obstacles in applications

- covariates or prognostic factors
- delayed responses
- multiple endpoints
- randomized allocation
- practical implementation
- statistical analysis

5. Bandit Processes

2). Practical implementation: RPW$(\alpha, \beta, \gamma; \delta)$

- deterministic: (Zelen, 1969)

the safety of prophylaxis with enoxaparin and dextran-70 in patients undergoing digestive surgery (Reiertsen et al, 1993, 94, 96, 97, 98)

- randomized: (Wei and Durham 1978)

urn model RPW$(\alpha, \beta, \gamma; \delta)$

ECMO trial: RPW(1, 1, 1, 1) (Bartlett et al 1985)

two anti-depression drug trials (Tamura et al 1994)

2). RPW$(\alpha, \beta, \gamma; \delta)$ - continued

Rosenberger (1996)

Hardwick et al (2001)

Ivanova (2003)

Wang and Prior (2003): $\delta = 2n + 1$

$\boxed{\text{5. Bandit Processes}}$

3). Delayed responses

- Motivation: survival trials

Eick (1988a, 1988b): geometric

Wang (2000, 2002): geometric

- unknown treatment X: survival times are geometric with unknown $\theta \in (0, 1)$

- the known treatment Y: survival times with a known expected value $k > 1$

- objective: maximize $W(\pi) = E_\pi(\Sigma_{i=1}^{\infty} \alpha_i Z_i)$

$Z_i = i^{th}$ patient's survival time under π

5. Bandit Processes

3). Delayed responses - model

- Bayesian approach: $\theta \sim \mu$ prior
- sufficient statistics: (s, f) on unknown
- posterior: $(s, f)\mu$, $(0, 0)\mu = \mu$
- posterior expected survival time: $E(X|(s, f)\mu)$
- state of the bandit: $((s, f)\mu, r, D)$
- r: size of the information bank
- $D = (\alpha_1, \alpha_2, \cdots)$: discount sequence
- optimality equation:

$$
\begin{aligned}
& V((s, f)\mu, r, D) \\
= & \max\{V^{(x)}((s, f)\mu, r, D), V^{(y)}((s, f)\mu, r, D)\}
\end{aligned}
$$

5. Bandit Processes

3). Delayed responses - optimal strategy

- advantage of treatment X over Y:

$$\Delta((s,f)\mu,r,D)$$
$$= V^{(x)}((s,f)\mu,r,D) - V^{(y)}((s,f)\mu,r,D)$$

- optimal strategy: treatment X optimal iff $\Delta((s,f)\mu,r,D) \geq 0$;

both optimal if $\Delta((s,f)\mu,r,D) = 0$

- *Condition A.* $\alpha_i \geq \Sigma_{j=i+1}^{\infty} \alpha_j$ for $i = 1, 2, \cdots$

- *Condition B.* μ is not concentrated at a single point, and $\mu\{(0,1)\} = 0$

$\boxed{\text{5. Bandit Processes}}$

3). Delayed responses - existence

THEOREM 1 *(Eick 1988)*

1. $\Delta((s,f)\mu,r,D)$ *is nonincreasing in* f *and* k *and nondecreasing in* s*. Also,* $\Delta((s,f)\mu,r+1,D) \geq \Delta((s,f+1)\mu,r,D)$.

2. For given f,r *and* k*, let* s^* *be such that* $\Delta((s^*,f)\mu,r,D)=0$*. Treatment* X *is optimal at the state* $((s,f)\mu,r,D)$ *iff* $s \geq s^*$*. Both are optimal if* $s=s^*$.

3. Let $D=(1,\alpha,\alpha^2,\cdots)$ *be geometric. If the known treatment* Y *is optimal at the state* $((s,f)\mu,0,D)$*, then it remains optimal for all subsequent patients.*

3). Delayed responses - structures

THEOREM 2 *(Wang 2000)*

1. $0 \leq s^* \leq s_1^*$, *and* $0 < s^* < s_1^*$ *under some conditions,* $E(X|(s_1^*, f)\mu) = k$.

s^* *is nondecreasing in both* f *and* k.

2. If $\Delta((s_n^*, f)\mu, 0, D_n) = 0$, *then*

$$0 \leq \cdots \leq s_n^* \leq \cdots \leq s_2^* \leq s_1^*,$$

and $s^* = \lim_{n\to\infty} s_n^*$ *exists, where* $D = (1, \alpha, \cdots)$, $D_n = (1, \alpha, \cdots, \alpha^{n-1}, 0, \cdots)$

3. If the known treatment is optimal at state $((s, f)\mu, 0, D_n)$, *then it remains optimal for the rest of the patients.*

$\boxed{\text{5. Bandit Processes}}$

3). Delayed responses - asymptotics

THEOREM 3 *(Wang 2002)*

1. If $\Delta((s_n^*(r,f),f)\mu,r,D)=0$, *then*

$$\lim_{r\to\infty}[\Delta((s,f)\mu,r,D)-\Delta((s,f)\mu,r,D_1)]=0,$$

$$s_1^*(\infty,f)=\cdots=s_n^*(\infty,f)=\cdots=s_1^*(0,f).$$

2. $\lim_{f\to\infty} s_n^*(r,f)=\infty$ *for any* r *and* n.

3. $\lim_{f\to\infty}\Delta((s_n^*(0,f),f)\mu,r,D_n)=0.$

5. Bandit Processes

3). Delayed responses - continuous model

Wang and Bickis (2003)

- treatment times: $T_1 \equiv 0, T_2, \cdots, \cdots$; $T_{n+1} - T_n \sim H(u)$, $\int_0^\infty u dH(u) < \infty$, $H(0) = 0$

- treatment 1: $X_{1,n} \sim F$, unknown

- treatment 2: $E(X_{2,n}) \equiv \lambda$

- Bayesian: $F \sim G \in \mathcal{D}(\mathcal{D})$

- $\mathcal{D}$: all distributions on $[0, \infty)$

- $\mathcal{D}(\mathcal{D})$: all distributions on $\mathcal{D}$

- possibly censored observation: (x, δ)

- information set at t: $\mathcal{H}(t)$, $\mathcal{H}(0) = \emptyset$

5. Bandit Processes

3). Delayed responses - continuous model

- strategy: $\pi(\mathcal{H}(t)) \in \{1, 2\}$

- survival times: $Z_n = X_{\pi(\mathcal{H}(t_n)),n}$

- discrete discount: $D = (\alpha_1, \alpha_2, \cdots)$, $\alpha_n \geq 0$, $\Sigma_{n=1}^{\infty} \alpha_n < \infty$

- continuous discount: $\beta(t)$, $\beta(t) > 0$, $\sum_{n=1}^{\infty} \alpha_n \beta_{n-1} < \infty$, $\beta_0 = \beta(0)$, $\beta_n = \int_0^{\infty} \beta(t) dH^{*n}(t)$, H^{*n}: convolution of H

- updated discounts: $D^{(n-1)} = (\alpha_n, \alpha_{n+1}, \cdots)$, $\beta^{t_n}(s) = \beta(t_n + s)$

- state at time t_n: $s_n = (G_n, t_n, D^{(n-1)}, \beta^{t_n})$

5. Bandit Processes

3). Delayed responses - continuous model

- initial state: $s = (G, 0, D, \beta)$

- objective: maximize

$$W(s, \lambda, \pi) = E_\pi \left(\sum_{n=1}^{\infty} \alpha_n \beta(T_n) Z_n | G \right)$$

- maximum: $V(s, \lambda) = \sup_\pi W(s, \lambda, \pi)$

- optimal strategy:

$$\Delta(s_n, \lambda) = V^{(1)}(s_n, \lambda) - V^{(2)}(s_n, \lambda)$$

where

$$V^{(i)}(s_n, \lambda) = \sup_{\pi \in \Pi^{(i)}} W(s_n, \lambda, \pi)$$

5. Bandit Processes

3). Delayed responses - continuous model

THEOREM 4 *For any* $s = (G, t, D, \beta)$, *there is a* $\Lambda(s)$ *such that* $\Delta(s, \Lambda(s)) = 0$.

So the unknown treatment is optimal at state s *iff* $\lambda \leq \Lambda(s)$ *and the known treatment is optimal iff* $\lambda \geq \Lambda(s)$.

THEOREM 5 *If* $\Delta(s_n, \Lambda(s_n)) = 0$ *at* $s_n = (G, 0, D_n, \beta)$, *then*

$$E(X|G) = \Lambda(s_1) \leq \cdots \leq \Lambda(s_n) \leq \cdots.$$

Also, $\lim_{n\to\infty} \Lambda(s_n) = \Lambda(s) < \infty$ *exists such that* $\Delta(s, \Lambda(s)) = 0$, $s = (G, 0, D, \beta)$.

5. Bandit Processes

4) Any number of arms (continuum)

- possible application: optimal dosing

- index set of treatments I: finite, countably infinite, compact

- responses: $X_{i,n} \sim F_i, n = 1, \cdots, i \in I$

- $F = (F_i, i \in I)$: $\sup\limits_{i \in I} \int_0^\infty x F_i(dx) < \infty$

- Markov decision process: $(\mathcal{S}, I, q, r, W)$

- objective: maximize

$$W(s_0, \pi) = E_\pi\left(\sum_{n=0}^{\infty} \beta^n r(s_n, i_n) | s_0\right)$$

4) Any number of arms (continuum)

THEOREM 6 *(Bickis and Wang 2003)*

1. *If I is finite, then there is an optimal stationary strategy.*

2. *If I is countable, then there is a stationary strategy which is ϵ-optimal.*

3. *If I is compact and F has a conjugate prior distribution, then there is an optimal stationary strategy.*

4. *In the presence of delayed responses, if I is compact and F has a conjugate prior distribution, then there exists an optimal deterministic strategy.*