Adaptive estimation: how to hit a moving target

Matthieu J.S. Brinkhuis Gunter Maris

Cito, Psychometric Research Center

University of Amsterdam

RCEC

RCEC Workshop on IRT and Educational Measurement
October 13, 2010

RCEC Research Center voor Examinering en Certificering

Cito now you know

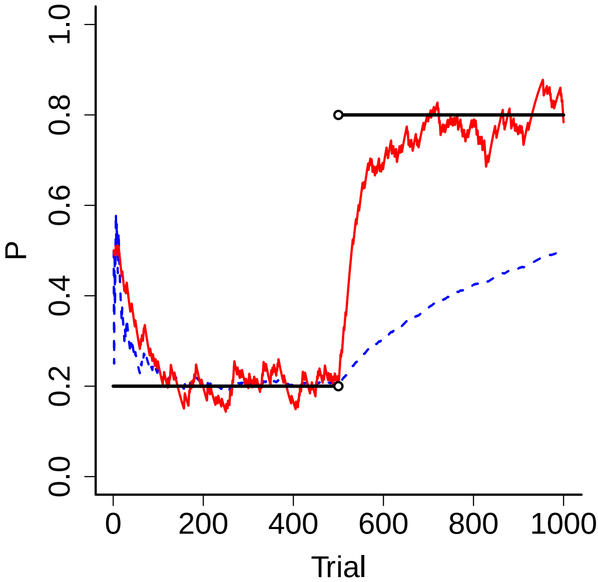
Adaptive estimation

In educational measurement parameters typically *change* with time

- Point estimation does not work for *changing* parameters
- Modeling is a straightforward approach, yet we often do not know *how* change occurs
- *Tracking* systems aim to follow development through time, rather than modeling it

Illustrate tracking with a simple coin tossing example

Flipping a biased coin



Two estimators

Two estimators where used in the coin flip illustration:

Average

$$\overline{X}_n = \frac{n-1}{n}\overline{X}_{n-1} + \frac{1}{n}X_n \quad (1)$$

Weights depend on n

Alternative estimator

$$\widehat{X}_n = (1-\alpha)\widehat{X}_{n-1} + \alpha X_n \quad (2)$$

Weights fixed

Alternative estimators

Several alternative estimators can be thought of:

Linear difference estimator (e.g. Batchelder & Bershad, 1979)

- The system is unbiased if provided with unbiased input (invariant distribution known)
- Single item response is not an unbaised estimator

Elo rating formula for continuous measurement (Elo, 1978)

Procedure for esimation of chess ability

$$R_n = R_o + K(W - W_e) \quad (3)$$

- The system is slightly biased
- The invariant distribution is unknown

An MCMC approach

A basic MCMC representation

$$\begin{array}{ccccccccc} & & \theta & & \theta & & \dots & & \theta \\ & & \downarrow & & \downarrow & & & & \downarrow \\ & & Y_1 & & Y_2 & & \dots & & Y_\infty \\ & & \downarrow & & \downarrow & & & & \downarrow \\ X_0 & \rightarrow & X_1 & \rightarrow & X_2 & \rightarrow & \dots & \rightarrow & X_\infty \end{array} \quad (4)$$

- New state determined from old state and stochastic innovation
- Convergence to invariant distribution X_∞ (typically complicated)

Our interest is in the distribution of X_∞ (known distribution with location parameter θ):

$$P(X_\infty \leq x) = F(x - \theta) \equiv F_\theta(x) \tag{5}$$

Two ways of using this distribution:

- If θ changes, the estimates follow (converge to a new invariant distribution)
- If θ remains stable, the chain converges with equal variance for all persons (CTT, useful for survey research)

$$\widehat{\theta}_t = \theta_t + \text{error} \tag{6}$$

Metropolis-Hastings

We use a Metropolis algorithm (Metropolis et al., 1953; Chib & Greenberg, 1995) to construct such a Markov chain

Characterization of the Metropolis algorithm

$$X_{t+1} \sim \begin{cases} X_t & \text{if } z_t = 0 \\ Y_t & \text{if } z_t = 1 \end{cases} \sim X_t \sim F_\theta(\cdot) \tag{7}$$

A Metropolis algorithm generates a Markov chain that involves three distribution:

1. f, the invariant distribution of states
2. g, the proposal distribution conditionally on the current state
3. π, the acceptance probability conditionally on both the current state and the proposal value

Adaptive estimation with Metropolis

Three distributions for Metropolis may be chosen freely if:

- The proposal distribution g is symmetric around the current state (Hastings, 1970)
- The acceptance distribution π meets the detailed balance condition

Since the distribution of states f should be *independent* of the items administered, we can only use item answers in g or π

$$X_{t+1} \sim \begin{cases} X_t & \text{if } z_t = 0 \\ Y_t & \text{if } z_t = 1 \end{cases} \sim X_t \sim F_\theta(\cdot) \tag{8}$$

Item responses as stochastic innovations

$$X_{t+1} \sim \begin{cases} X_t & \text{if } z_t = 0 \\ Y_t & \text{if } z_t = 1 \end{cases} \sim X_t \sim F_\theta(\cdot) \tag{9}$$

The invariant distribution f is Bernoulli with:

$$p = \frac{\exp(\theta)}{1 + \exp(\theta)} \tag{10}$$

If item responses Y are assumed to be generated by:

$$\Pr(Y_t = 1) = \frac{\exp(\theta - \delta_{I(t)})}{1 + \exp(\theta - \delta_{I(t)})} \quad (11)$$

and we choose the following acceptance probabilities:

$$\pi(1,0) = \min[1, \exp(\delta_{I(t)})] \quad (12)$$

and

$$\pi(0,1) = \min[1, \exp(-\delta_{I(t)})] \quad (13)$$

then

$$X_{t+1} \sim \begin{cases} X_t & \text{if } z_t = 0 \\ Y_t & \text{if } z_t = 1 \end{cases} \sim X_t \sim F_\theta(\cdot) \quad (14)$$

Simulations: single person

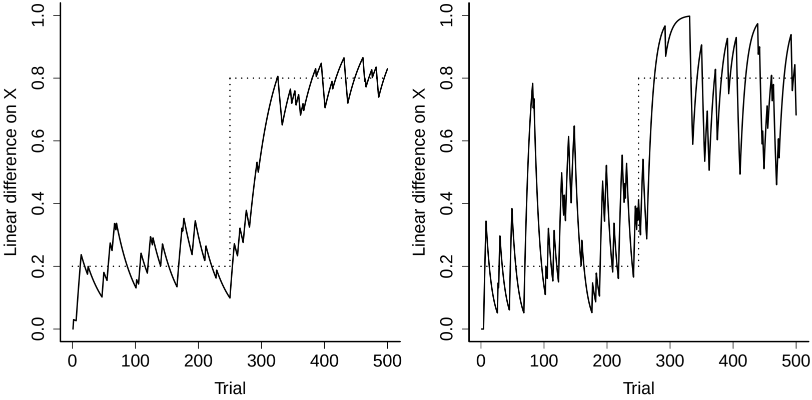


Figure: LD on a person, $\alpha = .03$ and $\alpha = .1$.

Simulations: group

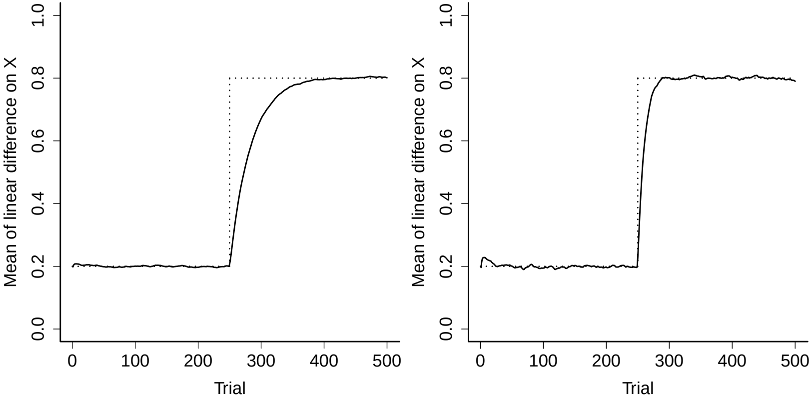


Figure: Mean LD, $\alpha = .03$ and $\alpha = .1$.

Item responses as acceptance variables

$$X_{t+1} \sim \begin{cases} X_t & \text{if } z_t = 0 \\ Y_t & \text{if } z_t = 1 \end{cases} \sim X_t \sim F_\theta(\cdot) \tag{15}$$

If we assume the following relation:

$$\pi(x,y) = \frac{f(y-\theta)}{f(x-\theta)+f(y-\theta)} \tag{16}$$

and assume a normal distribution for f we obtain that:

$$\pi(x,y) = \frac{\exp\left[\frac{y-x}{\sigma^2}\left(\theta - \frac{y+x}{2}\right)\right]}{1+\exp\left[\frac{y-x}{\sigma^2}\left(\theta - \frac{y+x}{2}\right)\right]} \tag{17}$$

here we recognize a 2PL model

Simulations

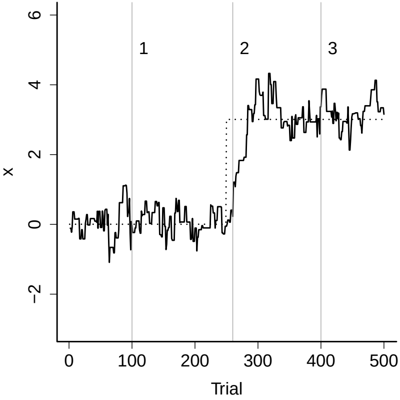


Figure: A changing parameter value (dotted line) and adaptive estimate (solild line).

Simulations (cont.)

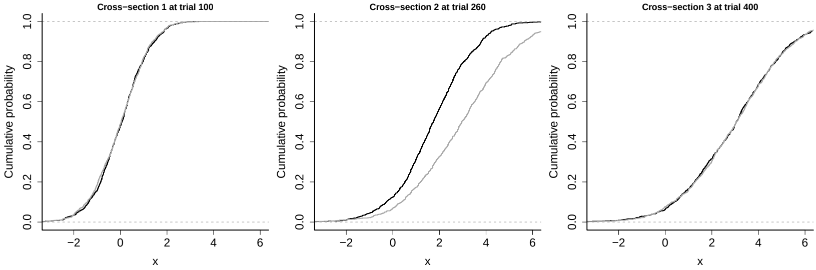


Figure: Cumulative empirical distributions of true abilities with error (grey) and adaptive estimates (black) at cross-section 1,2 and 3.

Discussion

Two simple solutions have been achieved with two main applications:

- Tracking individual ability
- Survey research

Some open problems remain:

- Goodness of fit
- Concurrent calibration

Thank you

References

Batchelder, W. H., & Bershad, N. J. (1979). The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology*, *19*(1), 39-60.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, *49*(4), 327–335.

Elo, A. E. (1978). *The rating of chess players, past and present*. London: B.T. Batsford, Ltd.

Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*(6), 1087–1092.

Algorithm proposal in pseudo-code

if $y_t = x_{t-1}$ then $x_t = x_{t-1}$
else
if $x_{t-1} = 1$ and $y_t = 0$ then $z \sim Bernoulli(\exp(\delta_{I(t)}))$
if $x_{t-1} = 0$ and $y_t = 1$ then $z \sim Bernoulli(\exp(-\delta_{I(t)}))$
if $z = 0$ then $x_t = y_t$
if $z = 1$ then $x_t = x_{t-1}$
$\theta_t = (1 - \alpha)\theta_{t-1} + \alpha x_t$

Algorithm acceptance probability in pseudo-code

$y \sim \mathcal{N}(x_t, \sigma)$
$a = (y - x_t)/\sigma^2$
$\delta = (y + x_t)/2$
candidate generates z_t
if $z_t = 1$ then $x_t = y$
if $z_t = 0$ then $x_t = x_{t-1}$