

Chapter 8

Tracking Ability: Defining Trackers for Measuring Educational Progress



Matthieu J. S. Brinkhuis and Gunter Maris

Abstract In measurements that extend over longer periods of time, we might expect all sorts of changes in model parameters, such as ability and item difficulty. We define trackers as instruments with specific properties to deal with such changing parameters. First, trackers should allow for estimating dynamically changing parameters, adapting to possible changes in ability or item difficulty. Second, if no change occurs for some time, trackers should provide unbiased estimates with a known error variance. Such trackers retain the strengths of both state space models and rating systems, while resolve some of their weaknesses. These properties are especially suitable for educational measurement applications such as tracking individual progress or any aggregate thereof, as in reporting survey research.

8.1 Introduction

In this chapter, we describe trackers and their uses in educational measurement. For now, we loosely define trackers as dynamic parameter estimates, adapting to possible changes in ability or item difficulty. Trackers can be especially useful in measurements that extend over a longer period of time at irregular time intervals, e.g., the continual measurement of abilities in computer adaptive practice (CAP) or computer adaptive learning (CAL) (Brinkhuis et al. 2018; Klinkenberg et al. 2011; Wauters et al. 2010; Veldkamp et al. 2011) or the monitoring of item difficulties in item banks (Brinkhuis et al. 2015). Many actors can be involved in the possible changes of these parameters, including the pupils themselves, their teachers, their parents, educational reforms, etc. Moreover, change in parameters, and the model

M. J. S. Brinkhuis (✉)
Department of Information and Computing Sciences,
Utrecht University, Utrecht, The Netherlands
e-mail: m.j.s.brinkhuis@uu.nl

G. Maris
ACTNext, Iowa City, IA, USA
e-mail: marisg@act.org

© The Author(s) 2019
B. P. Veldkamp and C. Sluijter (eds.), *Theoretical and Practical Advances in Computer-based Educational Measurement*, Methodology of Educational Measurement and Assessment, https://doi.org/10.1007/978-3-030-18480-3_8

itself, is especially likely if the outcomes of the measurements are used for feedback, as in assessment for learning (Black and Wiliam 2003; Bennett 2011; Wiliam 2011). Since feedback is provided to many actors in education, the result is a complex dynamical system, including all sorts of interactions.

The development of these parameters is not easily modeled due to these changes and feedback loops. Application of latent growth models (McArdle and Epstein 1987; Meredith and Tisak 1990; Hox 2002), change point estimation models (Hinkley 1970; Chib 1998; Visser et al. 2009) or other models that explicitly model the development of parameters is therefore not straightforward. Also state space models such as the Kalman filter (KF) (Kalman 1960; Welch and Bishop 1995; van Rijn 2008), or the more general particle filters (Arulampalam et al. 2002), include an explicit growth model and are therefore not ideal for following educational progress continually, as in CAP systems.

Historically, we see that in other fields where continual progress measurements take place, rating systems emerged. For example, in chess, rating systems were developed for the estimation of continually changing chess playing abilities, such as the widely used Elo rating system (ERS) (Elo 1978; Batchelder and Bershad 1979; Batchelder et al. 1992). Advantages of rating systems such as Elo's are that they are computationally light and do not assume a growth model. After each new measurement, the parameter estimates can be updated using the previous parameter estimate and the new observation, without having to take history into account, i.e., satisfying the Markov property, nor assuming a model for the development of the parameters. The ERS has found many uses, including applications in educational measurement (Klinkenberg et al. 2011; Wauters et al. 2010; Pelánek et al. 2017). Though there are many practical uses of rating systems, they lack certain desirable statistical properties such as convergence and unbiasedness (Brinkhuis and Maris 2009).

In tracking educational progress, we are interested in several properties from both state space models such as KFs and rating systems like the ERS, i.e., we like to dynamically track changes in the abilities of individuals or in item difficulties, without having to assume specific types of development. Moreover we require that, if ability is stable for some time, our tracker provides unbiased estimates with a known error variance. In this chapter, we will describe a tracker with these properties.

8.2 Methods

8.2.1 Formalizing a Tracker

We formalize the representation of a tracker in the scheme in (8.1), where we illustrate the development of someone's unknown true ability θ over time t , where each column represents a consecutive time point:

$$\begin{array}{ccccccc}
 \text{ability} & & \theta_1 & \rightarrow & \theta_2 & \rightarrow & \theta_3 & \rightarrow & \dots & \rightarrow & \theta_t \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\
 \text{responses} & & Y_1 & & Y_2 & & Y_3 & & \dots & & Y_t \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\
 \text{estimates} & X_0 & \rightarrow & X_1 & \rightarrow & X_2 & \rightarrow & X_3 & \rightarrow & \dots & \rightarrow & X_t
 \end{array} \tag{8.1}$$

The abilities θ are related by horizontal arrows, since we assume that one’s true ability at time point t is related at least to one’s ability at time point $t - 1$ and likely influenced by many other factors, which we leave out of this scheme. At time point t , scored responses Y_t are obtained using a single item, or a number of items. The ability estimate X_t depends only on the previous state X_{t-1} and the current item response Y_t , therefore satisfying the Markov property. The scheme in (8.1) represents Markov chains in general, including the ERS.

Since we are especially interested in the properties of unbiasedness and convergence, we present a more specific scheme in (8.2). Here, we assume for the moment that someone’s ability does *not* change, i.e., $\theta_t = \theta \forall t$, and we require X_∞ to have a known distribution, for example centered around the true ability θ with normal distributed error \mathcal{E} :

$$\begin{array}{ccccccc}
 \text{ability} & & \theta & \rightarrow & \theta & \rightarrow & \theta & \rightarrow & \dots & \rightarrow & \theta \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\
 \text{responses} & & Y_1 & & Y_2 & & Y_3 & & \dots & & Y_\infty \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\
 \text{estimates} & X_0 & \rightarrow & X_1 & \rightarrow & X_2 & \rightarrow & X_3 & \rightarrow & \dots & \rightarrow & X_\infty \sim \theta + \mathcal{E}
 \end{array} \tag{8.2}$$

We want to create a tracking algorithm that provides estimates X_t that adapt to changes in θ_t , but has a known distribution if θ_t is invariant for some t , as represented in (8.2). Such a tracking algorithm is similar to KFs in that its states have a known distribution (Arulampalam et al. 2002), and similar to the ERS, specifically Elo’s Current Rating Formula (Elo 1978), in that it continually adapts to changes in the underlying parameters without having to specify a growth model. An illustration of a simple tracker that conforms to this definition is given in Sect. 8.2.2, after which a proof of convergence is given in Sect. 8.2.3.

8.2.2 Example of a Tracker

We present a simple non-trivial case of a tracker that conforms to scheme (8.1), i.e., it dynamically adapts to change in the model parameters, and converges to a known error distribution if the scheme in (8.2) holds.

8.2.2.1 Coin Tossing Tracker

Consider the following coin tossing example.

$$\Pr(Y_i = 1|\theta) = \theta \quad (8.3)$$

where the probability of tossing head, i.e., $Y_i = 1$, is θ . If we consider a bandwidth of n sequential coin flips, then we simply define the sum score $X_+^{(n)}$ as follows:

$$X_+^{(n)} = \sum_{i=1}^n Y_i \sim \text{binom}(n, \theta) \quad (8.4)$$

Since (Y_1, \dots, Y_n) is independent of $\theta|X_+^{(n)}$, we can define an auxiliary variable Z using the sufficient statistic X_+ :

$$\Pr(Z = 1|X_+^{(n)} = x_+) = \frac{x_+}{n} = \Pr(Y_i = 1|X_+^{(n)}, \theta). \quad (8.5)$$

Using the auxiliary variable Z , which is the expected response given the sum score $X_+^{(n)}$, and the observed response Y , we readily find the following sequential update rule for $X_+^{(n)}$, which is denoted with the subscript t as an index for time:

$$X_{t+1}^{(n)} = X_t^{(n)} + Y_t - Z_t \sim \text{binom}(n, \theta) \quad (8.6)$$

which gives us the simplest non-trivial tracker $X_t^{(n)}/n$ for θ_t meeting our definition in (8.2).

We provide some illustrations to demonstrate the workings of the tracker in (8.6), using simulations.¹

8.2.2.2 Illustration of Convergence

First, we demonstrate the convergence of the sequential estimates of $X_t^{(n)}$ to the invariant distribution, $\text{binom}(n, \theta)$. As data, 1000 coin tosses are simulated with $\theta = .3$. Using the algorithm in (8.6) with $n = 30$, $X_t^{(n)}$ was sequentially estimated on the data and its distribution plotted in Fig. 8.1. As a reference, the theoretical density of the binomial distribution ($n = 30, \theta = .3$) was added. Clearly, this tracker nicely converged to the expected distribution as the two lines in Fig. 8.1 coincide. While the simulation used an invariant probability of the coin falling heads with $\theta = .3$, i.e., conforming to the scheme in (8.2), we can also simulate different changes to θ over time, conforming to the scheme in (8.1).

¹All simulation in this chapter are performed in R (R Core Team 2015).

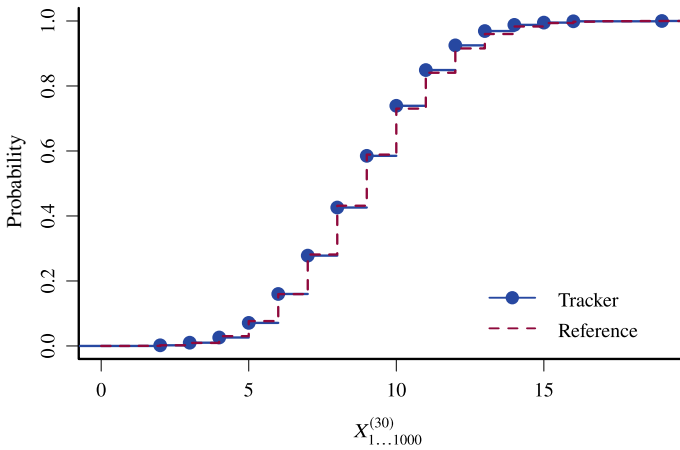


Fig. 8.1 Theoretical and empirical cumulative score distribution

8.2.2.3 Illustration of Tracking Smooth Growth

We simulate a scenario where θ smoothly changes over time t , i.e., we generate 1000 coin tosses with an increasing θ , and evaluate the development of the tracker in Fig. 8.2. Though θ is not stable at any time, it is clear that the tracker follows the development of θ quite closely with little lag. The step size n of the algorithm in (8.6) determines how fast the tracker can adapt to the changes in θ , where for this specific tracker a large n corresponds to a small step size and a small n to a large step size. Since θ is continually changing here, the tracker does not converge, but tracks the change rather well.

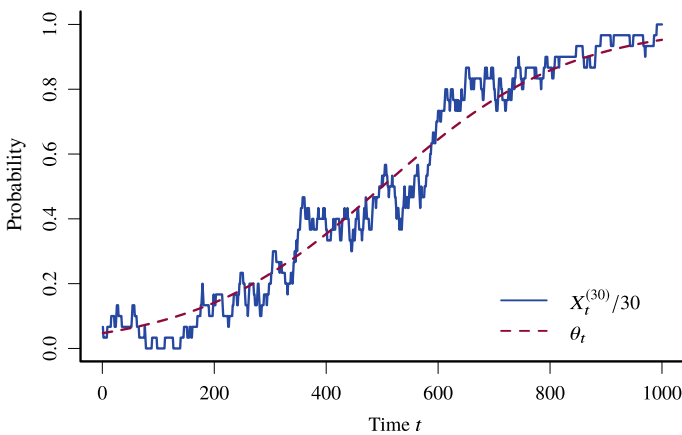


Fig. 8.2 Tracking smooth growth

8.2.2.4 Illustration of Tracking Sudden Changes

Next, we simulate a change point growth model where the probability of the coin falling heads changes from $\theta = .3$ to $\theta = .8$ at $t = 500$. The tracker is plotted in Fig. 8.3. Again, it can be seen that the tracker follows the development of θ closely. The tracker is always lagging, i.e., its development follows the development of θ with some delay depending on the step size n of the algorithm. This lag can be observed after the change point, and its size is related to the size of the change and step size of the algorithm.

8.2.2.5 Illustration of Varying Step Sizes

In Fig. 8.4 we illustrate the effect of varying the step size. A smooth development of θ is simulated, developing from about .1 to just over .8. Three different trackers are simulated using three different step sizes, $n = 10$, $n = 50$, and $n = 100$, where a small n is a large step size. It can be seen that the tracker with the most noise, having the largest step size and therefore the smallest n , adapts quickest to changes in ability θ , where the tracker with the smallest step size shows less noise, and therefore quite some lag. The step sizes are straightforward bias-variance trade-offs. Large step sizes allow for a quick adaption to possibly large changes in θ , at the cost of quite some variance if θ is stable. Small step sizes reduce this variance at the risk of introducing bias under a changing θ .

In these examples, this simplest non-trivial example of a tracker has demonstrated that it tracks the development of θ , cf. the scheme in (8.1), and converges to an invariant distribution if θ is stable, cf. the scheme in (8.2). We like to point out that though this example is simple, i.e., the properties of this tracker in the case of

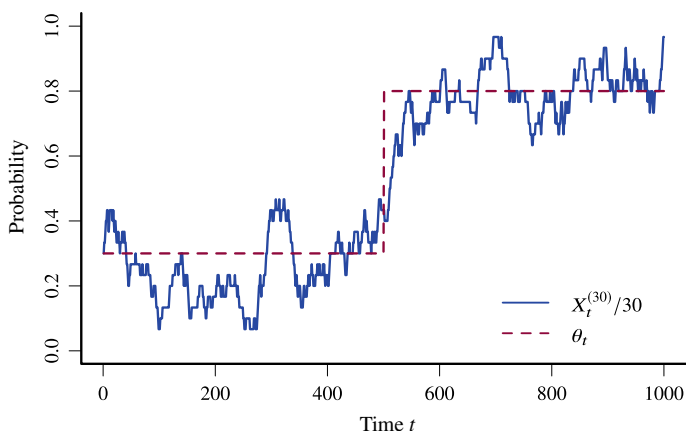


Fig. 8.3 Tracking a change point growth model

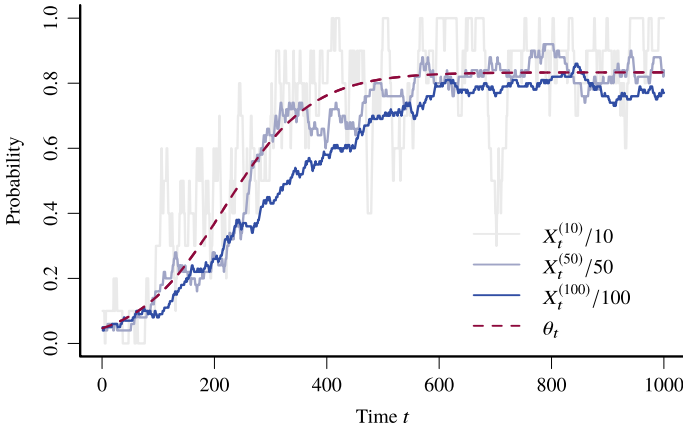


Fig. 8.4 Trackers with step sizes $n = 10$, $n = 50$ and $n = 100$

simple coin flips are the same as using a moving average, it is noteworthy that such trackers differ substantively from both maximum likelihood estimation (MLE) and Bayesian estimation techniques. The tracker estimates are continually adapting to changes in the model parameters, and convergence in distribution takes place while both the model parameter and the transition kernel are unchanging. This property of convergence under the scheme in (8.2) is generalized for all trackers in the following section.

8.2.3 Convergence in Kullback-Leibler Divergence

We provide some general proof of the convergence of Markov chains to an invariant distribution, given this distribution does not change between two time points, t and $t + 1$. We use the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951; Eggen 1999) to quantify the divergence between the current distribution f_t and the invariant distribution f_∞ .

Theorem 1 (Convergence in KL divergence) *If the invariant distribution $f_\infty(x)$ and the transition kernel $f_\infty(x|y)$ do not change between t and $t + 1$, the KL divergence between the current distribution and the invariant distribution decreases between two time points:*

$$\int_{\mathcal{R}} \ln \left(\frac{f_\infty(x)}{f_{t+1}(x)} \right) f_\infty(x) dx \leq \int_{\mathcal{R}} \ln \left(\frac{f_\infty(y)}{f_t(y)} \right) f_\infty(y) dy. \tag{8.7}$$

if

$$f_\infty(x) = \int_{\mathcal{R}} f_\infty(x|y) f_\infty(y) dy \tag{8.8}$$

and

$$f_{t+1}(x) = \int_{\mathcal{R}} f_{\infty}(x|y)f_t(y)dy \quad (8.9)$$

Proof Using Bayes' rule, we can rewrite (8.9) as follows:

$$f_{t+1}(x) = \int_{\mathcal{R}} \frac{f_{\infty}(y|x)f_{\infty}(x)}{f_{\infty}(y)}f_t(y)dy \quad (8.10)$$

and place $f_{\infty}(x)$ outside the integral:

$$\frac{f_{t+1}(x)}{f_{\infty}(x)} = \int_{\mathcal{R}} \frac{f_t(y)}{f_{\infty}(y)}f_{\infty}(y|x)dy. \quad (8.11)$$

Taking the logarithm and integrating with respect to $f_{\infty}(x)$ gives:

$$\int_{\mathcal{R}} \ln \left(\frac{f_{t+1}(x)}{f_{\infty}(x)} \right) f_{\infty}(x)dx = \int_{\mathcal{R}} \ln \left(\int_{\mathcal{R}} \frac{f_t(y)}{f_{\infty}(y)}f_{\infty}(y|x)dy \right) f_{\infty}(x)dx. \quad (8.12)$$

Using Jensen's inequality, we obtain:

$$\int_{\mathcal{R}} \ln \left(\int_{\mathcal{R}} \frac{f_t(y)}{f_{\infty}(y)}f_{\infty}(y|x)dy \right) f_{\infty}(x)dx \geq \int_{\mathcal{R}} \int_{\mathcal{R}} \ln \left(\frac{f_t(y)}{f_{\infty}(y)} \right) f_{\infty}(y|x)f_{\infty}(x)dydx \quad (8.13)$$

which we can use to simplify (8.12) into:

$$\int_{\mathcal{R}} \ln \left(\frac{f_{t+1}(x)}{f_{\infty}(x)} \right) f_{\infty}(x)dx \geq \int_{\mathcal{R}} \ln \left(\frac{f_t(y)}{f_{\infty}(y)} \right) f_{\infty}(y)dy. \quad (8.14)$$

Writing (8.14) as a KL divergence, we interchange numerators and denominators and therefore change the sign of the inequality:

$$\int_{\mathcal{R}} \ln \left(\frac{f_{\infty}(x)}{f_{t+1}(x)} \right) f_{\infty}(x)dx \leq \int_{\mathcal{R}} \ln \left(\frac{f_{\infty}(y)}{f_t(y)} \right) f_{\infty}(y)dy \quad (8.15)$$

which concludes our proof. \square

It was proven quite generally that trackers as described by (8.2) possesses an attractive quality. After every item response, the ability distribution of X_t monotonically converges in KL divergence to X_{∞} (Kullback and Leibler 1951). The KL divergence is a divergence measure between two distributions, in our case, the theoretical distribution of ability estimates X_t and the invariant distribution of estimates X_{∞} . If the KL divergence is small, the ability estimates are (almost) converged to the proper invariant distribution. Monotone convergence assures this divergence decreases with every new response under the conditions of (8.2).

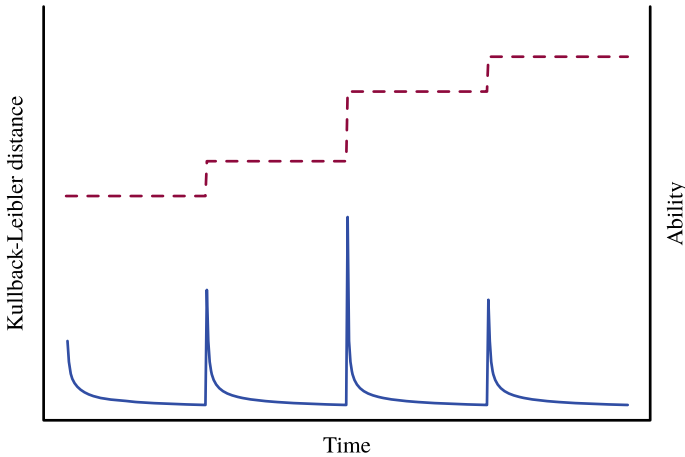


Fig. 8.5 Development of ability (dashed line) and Kullback-Leibler (KL) divergence (solid line) over time

8.2.3.1 Illustration of Development of Kullback-Leibler (KL) Divergence

In Fig. 8.5 we provide an illustration² how the KL divergence could develop over time. If ability θ is stable for some time, then changing, and then stable again, we can see how the KL divergence could decrease in times of stability and increase when ability changes.

We believe this illustration shows that the convergence property is suitable for use in the practice of educational measurement, where students mostly respond to sets of items, even if they are assessed frequently (Brinkhuis et al. 2018). The assumption here is that ability is stable during the relatively short time in which a student answers a set of items, and might change between the administrations of sets. Clearly, no convergence takes place if ability is continually changing (Sosnovsky et al. 2018).

8.2.4 Simulating Surveys

Section 8.2.2 provides some simulations to illustrate the invariant distribution of the simple tracker and to demonstrate how the estimates track individual development under several simulation conditions. One goal is to simulate and track the development of groups, as might be done in survey research.

We consider a simplified scenario where an entire population consisting of 100,000 persons would answer just 5 questions in a survey that is administered 4 times, for example to track educational progress of the entire population within

²This serves as illustration only, no actual KL divergences were calculated.

a year. Using the algorithm in (8.6), that compares to 5 flips of 100,000 uniquely biased coins for each survey. The probabilities of the coins falling heads changes for every survey, and are sampled from a beta distribution. The parameters of these beta distribution where $a = 5, 6\frac{2}{3}, 8\frac{1}{3}, 10$ and $b = 10$ for the 4 simulated surveys. These 4 beta distributions are plotted in Fig. 8.6 from left to right, using dashed lines. A very large step size ($n = 2$) was used for the algorithm, to allow the estimates \mathbf{X} to adapt quickly to the changes in θ . Since θ is sampled from a beta distribution, the estimates \mathbf{X} are beta-binomial distributed. Using MLE, the two parameters of the beta distribution of θ were estimated for each of the 4 administrations, and these estimated beta distributions are plotted in Fig. 8.6. The graph demonstrates that it is possible to accurately track the development of an entire population by administrating a limited amount of items to all individuals.

Note that this scenario is quite different from a more traditional sampling approach where many items are administered to complex samples of individuals. If the total number of responses would be kept equal, for tracking the development of the entire population it is beneficial to administer a few questions to many individuals. On the contrary, for tracking individual development, it is beneficial to administer many items to few individuals. For example, while the information in the 5 items per survey described above is too limited for tracking individual growth, especially in considering progress that is made between surveys, it is sufficient for tracking the population parameters. Though trackers can both be used for tracking individual progress or progress of the population, the preferred design of data collection depends on the desired level of inference.

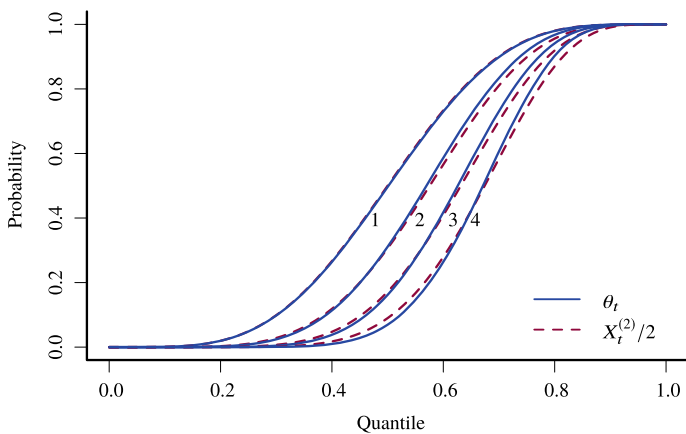


Fig. 8.6 Trackers with step sizes $n = 10, n = 50$ and $n = 100$

8.3 Discussion

In this chapter, trackers and their possible uses in educational measurement have been described. Trackers are defined as dynamic parameter estimates with specific properties. These trackers combine some of the properties of the ERS and state space models as KFs, which both have strengths and weaknesses. The ERS is a feasible method for dealing with data with changing model parameter, i.e., ratings. It is simple, provides real-time results, and requires no assumptions on the types of growth. However, it lacks a proper distribution of estimates, and therefore no statistics can be used on the estimates, e.g., to test for change, or to track any aggregate of estimates. KFs, on the other hand, do assume specific distributions of estimates, but need specified growth models, which are not readily available in many educational measurement applications.

Trackers should be able to adapt to changes in both model parameters and the transition kernel, cf. the scheme in (8.1). In addition, we require that the estimates converge in distribution if the model is invariant, cf. scheme (8.2). A simple example of a tracker conforming to this definition has been introduced in (8.6), with a transition kernel that creates a Markov chain with a binomial invariant distribution. A well-known technique for obtaining a transition kernel that creates a Markov chain with a specified distribution is called the Metropolis algorithm (Metropolis et al. 1953; Hastings 1970; Chib and Greenberg 1995). The Metropolis algorithm can be used to create a transition kernel that satisfies (8.2). The general proof that such Markov chains monotonically converge to their invariant distribution has been provided in Theorem 1.

While the simple binomial example might not have much practical use directly, other trackers can be developed to provide estimates with a known error distribution, for example an ability estimate X which is distributed $\mathcal{N}(\theta, \sigma^2)$. Two simple examples of such trackers are presented in Brinkhuis and Maris (2010). Such estimates could directly be used in other statistical analyses since the magnitude of the error does not depend on the ability level itself. These assumptions compare directly to the assumptions of classical test theory, where an observed score equals the sum of the true score and an uncorrelated error. Another simple application of using these estimates directly is to look at empirical cumulative distributions of ability estimates.

Trackers as defined in this chapter retain the strengths of both state space models and ratings systems, and resolve some of their weaknesses. Their properties are suitable for, among others, applications in educational measurement in either tracking individual progress or any aggregate thereof, e.g., classes or schools, or the performance of the entire population as in survey research. The algorithms remain relatively simple and light-weight, and therefore allow to provide real-time results even in large applications. They are unique in that they continually adapt to a new transition kernel and converge in distribution if there is an invariant distribution, which is quite different from both MLE and Bayesian estimation techniques.

Acknowledgements An earlier version of this manuscript is part of Brinkhuis (2014).

References

- Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174–188. <https://doi.org/10.1109/78.978374>.
- Batchelder, W. H., & Bershadt, N. J. (1979). The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology*, 19(1), 39–60. [https://doi.org/10.1016/0022-2496\(79\)90004-X](https://doi.org/10.1016/0022-2496(79)90004-X).
- Batchelder, W. H., Bershadt, N. J., & Simpson, R. S. (1992). Dynamic paired-comparison scaling. *Journal of Mathematical Psychology*, 36, 185–212. [https://doi.org/10.1016/0022-2496\(92\)90036-7](https://doi.org/10.1016/0022-2496(92)90036-7).
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- Black, P., & Wiliam, D. (2003). “In praise of educational research”: Formative assessment. *British Educational Research Journal*, 29(5), 623–637. <https://doi.org/10.1080/0141192032000133721>.
- Brinkhuis, M. J. S. (2014). *Tracking educational progress* (Ph.D. Thesis). University of Amsterdam. <http://hdl.handle.net/11245/1.433219>.
- Brinkhuis, M. J. S., Bakker, M., & Maris, G. (2015). Filtering data for detecting differential development. *Journal of Educational Measurement*, 52(3), 319–338. <https://doi.org/10.1111/jedm.12078>.
- Brinkhuis, M. J. S., & Maris, G. (2009). *Dynamic parameter estimation in student monitoring systems* (Measurement and Research Department Reports 09-01). Arnhem: Cito. <https://www.researchgate.net/publication/242357963>.
- Brinkhuis, M. J. S., & Maris, G. (2010). *Adaptive estimation: How to hit a moving target* (Measurement and Research Department Reports 10-01). Arnhem: Cito. <https://www.cito.nl/kennis-en-innovatie/kennisbank/p207-adaptive-estimation-how-to-hit-a-moving-target>.
- Brinkhuis, M. J. S., Savi, A. O., Coomans, F., Hofman, A. D., van der Maas, H. L. J., & Maris, G. (2018). Learning as it happens: A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, 5(2), 29–46. <https://doi.org/10.18608/jla.2018.52.3>.
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2), 221–241. [https://doi.org/10.1016/S0304-4076\(97\)00115-2](https://doi.org/10.1016/S0304-4076(97)00115-2).
- Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327–335. <https://doi.org/10.2307/2684568>.
- Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, 23(3), 249–261. <https://doi.org/10.1177/01466219922031365>.
- 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. <https://doi.org/10.1093/biomet/57.1.97>.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1), 1–17. <https://doi.org/10.1093/biomet/57.1.1>.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. New Jersey: Lawrence Erlbaum Associates Inc.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D), 35–45.
- Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. J. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <http://www.jstor.org/stable/2236703>.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58(1), 110–133. <https://doi.org/10.2307/1130295>.

- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122. <https://doi.org/10.1007/BF02294746>.
- 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. <https://doi.org/10.1063/1.1699114>.
- Pelánek, R., Papoušek, J., Říhák, J., Stanislav, V., & Nižnan, J. (2017). Elo-based learner modeling for the adaptive practice of facts. *User Modeling and User-Adapted Interaction*, 27(1), 89–118. <https://doi.org/10.1007/s11257-016-9185-7>.
- R Core Team. (2015). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing, Vienna, Austria*. <http://www.R-project.org/>.
- Sosnovsky, S., Müter, L., Valkenier, M., Brinkhuis, M., & Hofman, A. (2018). Detection of student modelling anomalies. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning* (pp. 531–536). Berlin: Springer. https://doi.org/10.1007/978-3-319-98572-5_41.
- van Rijn, P. W. (2008). *Categorical time series in psychological measurement* (Ph.D. Thesis). University of Amsterdam, Amsterdam, Netherlands. <http://dare.uva.nl/record/270555>.
- Veldkamp, B. P., Matteucci, M., & Eggen, T. J. H. M. (2011). Computerized adaptive testing in computer assisted learning? In S. De Wannemacker, G. Clarebout, & P. De Causmaecker (Eds.), *Interdisciplinary approaches to adaptive learning, communications in computer and information science* (Vol. 126, pp. 28–39). Berlin: Springer. https://doi.org/10.1007/978-3-642-20074-8_3.
- Visser, I., Raijmakers, M. E. J., & van der Maas, H. L. J. (2009). Hidden Markov models for individual time series. In J. Valsiner, P. C. M. Molenaar, M. C. Lyra, & N. Chaudhary (Eds.), *Dynamic process methodology in the social and developmental sciences* (Chap. 13, pp. 269–289). New York : Springer. https://doi.org/10.1007/978-0-387-95922-1_13.
- Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning*, 26(6), 549–562. <https://doi.org/10.1111/j.1365-2729.2010.00368.x>.
- Welch, G., & Bishop, G. (1995). *An introduction to the Kalman filter* (Technical Report TR 95-041). Chapel Hill, NC, USA: Department of Computer Science, University of North Carolina at Chapel Hill. http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf. Updated July 24, 2006.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

