

Estimating the prevalence and incidence of a state behavior: Models for interval
recording data and a novel observation system

James E. Pustejovsky and Daniel M. Swan

The University of Texas at Austin

Author Note

James E. Pustejovsky, Department of Educational Psychology, University of Texas at Austin. Daniel M. Swan, Department of Educational Psychology, University of Texas at Austin.

Address correspondence to James E. Pustejovsky, Department of Educational Psychology, University of Texas at Austin, 1 University Station D5800, Austin, TX 78712. Email: pusto@austin.utexas.edu.

Abstract

Data based on direct observation of behavior are used widely in many areas of educational and psychological research, particularly in applied research areas such as the treatment of behavioral disorders. A number of different methods are used to record data during direct observation, including continuous recording, momentary time sampling (MTS), partial interval recording (PIR), and whole interval recording (WIR). Among these methods, PIR and WIR have long been recognized as problematic because, as typically reported, the mean of such data measures neither the prevalence nor the incidence of the observed behavior. Though the problems with these methods have long been recognized, little research has examined methods of analyzing interval recording data other than simply taking the mean. This paper proposes a Alternating Poisson Process model for interval recording data that permits estimation of both prevalence and incidence via maximum likelihood or penalized maximum likelihood methods. The paper also describes a novel observation recording method that involve combinations of MTS, PIR, and WIR and that provides considerably more efficient estimators of prevalence and incidence.

Keywords: behavioral observation; interval recording; alternating Poisson process; Markov chain

Estimating the prevalence and incidence of a state behavior: Models for interval recording data and a novel observation system

Measurements derived from systematic, direct observation of human behavior are used in many areas of psychological and educational research. For example, direct observation of student classroom behavior is a primary component of several existing instruments for screening and diagnosis of emotional and behavioral problems (Volpe, DiPerna, Hintze, & Shapiro, 2005); direct observation of childrens' challenging behavior in home settings has been employed to collect pre- and post-test measures in randomized trials of behavioral interventions (e.g., Durand, Hieneman, Clarke, Wang, & Rinaldi, 2012); and direct observation of infant-parent interaction patterns is employed in studies of child development (Mann, Ten Have, Plunkett, & Meisels, 1991) and cross-cultural differences (Bornstein, 2002). Direct observation also plays a prominent role in single-case research, where it is used to assess individual responses to intervention by measuring changes in behavior over time (Kazdin, 2011).

Systematic direct observation procedures require that the behavior of interest have a clear operational definition, so that its occurrence or absence can be judged at a given point in time. In forming such an operational definition, is useful to distinguish between behaviors that are events, where each occurrence is of negligible duration, versus behaviors that are states, where individual episodes of behavior have positive duration (J. Altmann, 1974). The primary characteristic of an event behavior is its incidence, or frequency of occurrence per time unit. In contrast, a state behavior has two primary characteristics: incidence, which is the frequency (per unit time) with which new episodes of the behavior begin and prevalence, which is the proportion of time that the behavior occurs. Given an operationally defined behavior, measurements of its characteristics are obtained by recording data while observing the behavior (either in person, or by video-recording) for a specified length of time.

There are several different procedures for recording data during direct observation,

varying in ease of implementation, the level of detail in the resulting data, and the aspect of behavior to which the resulting measurement corresponds (for surveys of major recording procedures, see J. Altmann, 1974; Ayres & Gast, 2010; Hartmann & Wood, 1990; Primavera, Allison, & Alfonso, 1996). The most intensive procedure is continuous recording (sometimes called duration recording or real-time recording), in which the observer records the time at which each behavioral episode begins and ends. Data from continuous recording is very rich, in that it permits direct estimation of prevalence, incidence, and further aspects of the pattern of behavior; continuous recording data can also be subjected to more sophisticated forms of modeling (e.g., Bakeman & Quera, 2011; Haccou & Meelis, 1992). However, less effort-intensive data collection methods are often required, particularly for use in clinical and applied research settings.

Other commonly used systems for collecting behavioral data do not capture a complete record of the behavior during an observation session, but rather involve making observations only intermittently. Among intermittent recording systems, the three main procedures are momentary time sampling, partial interval recording, and whole interval recording. In all three methods, an observation session is divided into a fixed number of equally spaced intervals, of perhaps 10 or 15 s in length, and a binary data-point is recorded for each interval. The systems differ only in the rule for scoring each interval. Using momentary time sampling (MTS), an interval is scored as a one if a behavioral event is happening during the final moment of the interval (and is otherwise scored as a zero). Using partial interval recording (PIR, also known as one-zero sampling, modified frequency sampling, or Hansen sampling), an interval is scored as a one if the behavior occurs at any point during the interval. Using whole interval recording (WIR), an interval is scored as a one only if the behavior occurs for the entire duration of the interval. In some PIR and WIR systems, a small length of time is left between each interval so that the observer does not have to maintain continuous attention.

In many applications, the interval-by-interval data generated by these recording

systems is summarized by the proportion of intervals scored as a one. Often, only this summary proportion is used for later analysis, where it is interpreted as a measure of prevalence. However, whether it is reasonable to reduce the data to a summary proportion depends on which recording system is used to collect the data.

Under quite general modeling assumptions, the proportion of MTS intervals is an unbiased estimate of prevalence (Rogosa & Ghandour, 1991). Thus, reducing the data to the summary proportion is entirely reasonable if the investigator's interest is solely in the prevalence of the behavior. The drawback of doing so is that simple summaries of MTS data do not provide any measure of the behavior's incidence. Even if not of substantive interest, an estimates of incidence is needed in order to assess the magnitude of measurement error in the prevalence estimate.

Brown, Solomon, and Stephens (1977, see also Griffin & Adams, 1983) described a method for estimating both prevalence and incidence from MTS data. Their approach was to first posit a stochastic process for the underlying pattern of behavior as perceived by the observer, or what is often termed the behavior stream. The particular model they considered was an Alternating Poisson Process, which is a simple, two-state continuous time Markov chain where transitions between states follow exponential distributions. Brown and colleagues showed that if the behavior stream is generated by such a process, then interval-by-interval MTS scores follow a discrete-time Markov chain, from which closed-form expressions for the maximum likelihood estimators of prevalence and incidence can be derived.

Unlike MTS, PIR and WIR systems do not produce clearly interpretable summary measurements. Rather, the PIR summary proportion systematically over-estimates prevalence and the WIR summary proportion systematically under-estimates prevalence; in both cases, the extent of the bias depends on characteristics of the behavior as well as operational features of the recording system (Kraemer, 1979; Rogosa & Ghandour, 1991), making the construct interpretation of such data quite difficult. Consequently,

methodologists have long argued against the use of PIR and WIR systems (J. Altmann, 1974; Lane & Ledford, 2014; Mann et al., 1991, cf.). Despite such objections, the systems remain in common use, particularly as part of behavioral time series designs and single-case research (Lane & Ledford, 2014; Mudford, Taylor, & Martin, 2009; Rapp et al., 2007).

Little previous research has considered methods of analyzing PIR and WIR data beyond using the summary proportion. For PIR data, S. A. Altmann and Wagner (1970) proposed a transformation of the summary proportion as an estimate of incidence, motivated by a model in which behavioral episodes follow a Poisson process. While this model applies well to event behaviors, it is not a suitable description of state behaviors, where individual episodes have non-negligible duration. Suen and Ary (1986, 1989) have proposed a method for obtaining estimates of prevalence and incidence from PIR data, provided that the behavior stream conforms to certain conditions. However, their proposed procedure is not motivated by any explicit data-generating process, and later simulation studies reported that the method produces badly biased estimates (Rogosa & Ghandour, 1991, sec. 5.2). Pustejovsky and Swan (2014) proposed several methods for bounding the bias of the PIR summary proportion as an estimate of prevalence, based on various prior assumptions about the behavior stream. These methods are useful for analysis of summarized PIR data, as would be available from a published single-case study, but are not full models of the data-generating process. Other methods of analysis, involving fully specified data-generating models for the interval-by-interval scores, are therefore of interest.

This paper examines models for PIR and WIR data, from which principled estimates of prevalence and incidence can be obtained. Following Brown et al. (1977), we use an Alternating Poisson Process for the underlying behavior stream to derive a model for the interval-by-interval scores. Under this model, maximum likelihood estimates for prevalence and incidence can be obtained using conventional numerical

techniques (although they do not have closed-form expression). To remedy some problems with the maximum likelihood estimators, we introduce penalized likelihood estimators that have better operating characteristics and that can be tailored to express prior information about the behavioral parameters. The final section of the paper describes a novel procedure for intermittent recording of a behavior that entails combining MTS and interval recording methods and that can be used to obtain more efficient estimators of prevalence and incidence.

We use a parametric bootstrap procedure for obtaining interval estimates.

Alternating Poisson Process models

The Alternating Poisson Process is a stochastic model that can be used to describe a stream of behavior, as it is perceived in time. The model applies to state behaviors, where the behavior is either occurring or not occurring at any given point in time and where each episode of behavior has non-negligible duration. The stream of a state behavior can be described in terms of two components: sequentially ordered, non-overlapping episodes of behavior, which we will call event durations, and spans of time in between episodes, which we will call interim times. Let $\{Z(t), 0 \leq t\}$ denote the state of the behavior stream over the course of an observation session, where $Z(t) = 1$ indicates that an event is occurring at time t and $Z(t) = 0$ otherwise.

The Alternating Poisson Process makes several further assumptions. Specifically, it is assumed that event durations and interim times are mutually independent, random quantities, that the event durations follow an exponential distribution with mean $\mu > 0$, and that the interim times follow an exponential distribution with mean $\lambda > 0$. Under the model, the prevalence of the behavior is equal to the ratio of μ to the sum of μ and λ and the incidence of the behavior is equal to the reciprocal of the sum of μ and λ . We will denote prevalence by ϕ , where $0 < \phi < 1$, and incidence by ζ , where $\zeta > 0$. Finally, it is assumed that the process is in equilibrium, with $\Pr(Y(0) = 1) = \phi$. This assumption implies that there is a constant marginal probability of observing an event at any given point in time.

The Alternating Poisson Process is a special case of a continuous time Markov chain, and thus has the Markov property that the future evolution of the behavior depends only on the current state, but not on the past history of the behavior. More precisely, the probability that a behavior will be occurring t seconds into the future is independent of the state of the behavior for $0 \leq r < s$:

$$\Pr[Z(s+t) = 1 | Z(s) = a, Z(r) : 0 \leq r < s] = \Pr[Z(s+t) = 1 | Z(s) = a] \quad (1)$$

for $a = 0, 1$ and $s, t \geq 0$ (Kulkarni, 2010, Thm. 6.1). The assumption that the process is in equilibrium further implies that the probability that a behavior will be occurring t seconds into the future does not depend on the current time, i.e.,

$$\Pr[Z(s+t) = 1 | Z(s) = a] = \Pr[Z(t) = 1 | Z(0) = a]. \quad (2)$$

Let $p_a(t)$ denote the conditional probability that an event will be occurring t seconds into the future, given that the behavior is currently in state a , for $a = 0, 1$. These conditional probabilities can be expressed as follows:

$$\begin{aligned} p_0(t) &= \Pr(Z(t) = 1 | Z(0) = 0) = \phi \left[1 - \exp \left(\frac{-t\zeta}{\phi(1-\phi)} \right) \right] \\ p_1(t) &= \Pr(Z(t) = 1 | Z(0) = 1) = \phi + (1-\phi) \exp \left(\frac{-t\zeta}{\phi(1-\phi)} \right) \end{aligned} \quad (3)$$

(Kulkarni, 2010, Eq. 6.17).

Momentary Time Sampling

Consider observing a behavior stream generated by the Alternating Poisson Process and recording observations using momentary time sampling with $K+1$ recording times, equally spaced at intervals of length c . Denote the recorded data by the sequence of binary indicator variables X_0, X_1, \dots, X_K . The MTS interval data are a record of the state of the behavior stream process at fixed moments in time: $X_k = Z(ck)$ for $k = 0, \dots, K$.

Brown et al. (1977) demonstrated that MTS data follow a two-state, discrete-time Markov chain process with transition probabilities $\Pr(X_k = 1 | X_{k-1} = a) = p_a(c)$ and

$Pr(X_k = 0|X_{k-1} = a) = 1 - p_a(c)$ for $a = 0, 1$. Therefore, sufficient statistics for the process are given by the table counting the number of transitions with

$(X_{k-1} = a, X_k = b)$ for $a, b = 0, 1$ and $k = 1, \dots, K$; let $n_{ab} = \sum_{k=1}^K I(X_{k-1} = a, X_k = b)$.

Conditioning on X_0 , the log-likelihood of MTS data is then given by

$$\begin{aligned} l_{MTS}(\phi, \zeta) = & n_{01} \log \phi + n_{10} \log (1 - \phi) \\ & + (n_{01} + n_{10}) \log \left[1 - \exp \left(\frac{-\zeta c}{\phi(1 - \phi)} \right) \right] \\ & + n_{00} \log \left[1 - \phi + \phi \exp \left(\frac{-\zeta c}{\phi(1 - \phi)} \right) \right] \\ & + n_{11} \log \left[\phi + (1 - \phi) \exp \left(\frac{-\zeta c}{\phi(1 - \phi)} \right) \right]. \end{aligned} \quad (4)$$

Brown et al. (1977) provided closed-form expressions for the maximum likelihood estimates (MLEs) for ϕ and ζ based on this model. Let $\hat{p}_0 = n_{01}/(n_{00} + n_{01})$ and $\hat{p}_1 = n_{11}/(n_{10} + n_{11})$. The MLE for ζ exists only when $\hat{p}_0 < \hat{p}_1$. When this condition holds, the MLEs for ϕ and ζ are given by

$$\hat{\phi}_{MTS} = \frac{\hat{p}_0}{\hat{p}_0 + 1 - \hat{p}_1} \quad \text{and} \quad \hat{\zeta}_{MTS} = \frac{-\hat{p}_0(1 - \hat{p}_1) \log(\hat{p}_1 - \hat{p}_0)}{c(\hat{p}_0 + 1 - \hat{p}_1)^2}. \quad (5)$$

The probability that the MLEs are undefined or fall outside of the parameter space is not trivial, even when K is relatively large. In order for the estimates to fall

Table 1

Proportion of 2000 simulated MTS samples ($K = 40$) in which $0 < \hat{\phi}_{MTS} < 1$ and $0 < \hat{\zeta}_{MTS} < \infty$.

	$\zeta = 0.02$	0.05	0.1	0.2	0.25	0.4	0.5
$\phi = 0.1$	0.43	0.63	0.64	0.45	0.37	0.29	0.26
0.2	0.45	0.80	0.90	0.84	0.79	0.59	0.49
0.3	0.46	0.84	0.96	0.96	0.94	0.73	0.64
0.4	0.43	0.87	0.98	0.99	0.97	0.82	0.71
0.5	0.45	0.88	0.99	1.00	0.98	0.84	0.71

strictly within the parameter space, both a 0-1 transition and a 1-0 transition must be observed, so that $\hat{p}_0 > 0$, $\hat{p}_1 < 1$. Table 1 reports the proportion of 2000 simulated samples in which $0 < \hat{\phi}_{MTS} < 1$ and $0 < \hat{\zeta}_{MTS} < \infty$, with $K = 40$ and ζ scaled in terms of the interval length. Values of $\phi > 0.5$ are omitted because the behavior of the MTS estimators is symmetric about $\phi = 0.5$. The proportion of estimates falling within the parameter space decreases as prevalence becomes more extreme and as incidence becomes very infrequent or very frequent. The high proportion of samples in which the estimate of incidence is undefined represents a drawback to the use of maximum likelihood based on MTS data.

Partial Interval Recording

Consider observing a behavior stream generated by the Alternating Poisson Process and recording observations using partial interval recording. Suppose that one observes K intervals, where each interval includes c seconds of active observation time followed by d seconds of recording time. Let time $t_k = (k - 1)(c + d)$ denote the beginning of interval k . Let U_k indicate the PIR score from interval k , corresponding to the time from t_k to $t_k + c$. Following the PIR system, $U_k = 1$ if the behavior occurs at any point during the active portion of interval, and $U_k = 0$ otherwise. In terms of the behavior stream process,

$$U_k = I \left[0 < \int_0^c Z(t_k + s) ds \right] \quad (6)$$

for $k = 1, \dots, K$, where \int_0^c denote the definite integral over the half-open interval $[0, c)$.

Under the assumptions of the Alternating Poisson Process, the joint distribution of U_1, \dots, U_K can be derived as follows. Let $\psi_k, k = 2, \dots, K$ denote the probability that the behavior is occurring at time $t_k = (k - 1)(c + d)$, given the partial interval record up to that time. Let $\psi_1 = \phi$, which follows from the assumption that the process is in

equilibrium. We show in Appendix A that

$$\begin{aligned}\psi_k &= \Pr[Z(t_k) = 1 | U_1, \dots, U_{k-1}] \\ &= \left[\frac{\psi_{k-1}p_1(c+d) + (1 - \psi_{k-1})[p_0(c+d) - p_0(d)\exp(\frac{-\zeta c}{1-\phi})]}{1 - (1 - \psi_{k-1})\exp(\frac{-\zeta c}{1-\phi})} \right]^{u_{k-1}} [p_0(d)]^{(1-u_{k-1})}.\end{aligned}\quad (7)$$

Note that $Z(t_k) = 1$ implies that $U_k = 1$ with certainty, while

$$\Pr(U_k = 1 | Z(t_k) = 0) = 1 - \exp\left(\frac{-\zeta c}{1-\phi}\right).$$

It follows from the Markov property of the Alternating Poisson Process that

$$\begin{aligned}\Pr(U_k = 1 | U_1, \dots, U_{k-1}) &= \psi_k \Pr(U_k = 1 | Y(t_k) = 1) + (1 - \psi_k) \Pr(U_k = 1 | Y(t_k) = 0) \\ &= 1 - (1 - \psi_k) \exp\left(\frac{-\zeta c}{1-\phi}\right).\end{aligned}$$

The joint distribution of U_1, \dots, U_K can therefore be expressed as

$$\begin{aligned}\Pr(U_1 = u_1, \dots, U_K = u_K) &= \Pr(U_1 = u_1) \prod_{k=2}^K \Pr(U_k = u_k | U_1, \dots, U_{k-1}) \\ &= \prod_{k=1}^K \left[1 - (1 - \psi_k) \exp\left(\frac{-\zeta c}{1-\phi}\right) \right]^{u_k} \left[(1 - \psi_k) \exp\left(\frac{-\zeta c}{1-\phi}\right) \right]^{(1-u_k)}.\end{aligned}$$

The log-likelihood of ϕ and ζ , given observed PIR data u_1, \dots, u_K , is

$$l_{PIR}(\phi, \zeta) = \sum_{k=1}^K u_k \ln \left[1 - (1 - \psi_k) \exp\left(\frac{-\zeta c}{1-\phi}\right) \right] + (1 - u_k) \left[\ln(1 - \psi_k) - \frac{\zeta c}{1-\phi} \right]. \quad (8)$$

MLEs $\hat{\phi}_{PIR}, \hat{\zeta}_{PIR}$ are obtained by maximizing l_{PIR} using numerical techniques. Because the conditional probabilities ψ_1, \dots, ψ_K are defined recursively, it is cumbersome and computationally expensive to evaluate the score function corresponding to this likelihood. The simulation study reported in the next section therefore uses the Nelder-Mead algorithm (?), which does not require evaluation of the score function.

Just as with MTS data, MLEs based on PIR data do not always fall within the parameter space. Table 2 reports the proportion of 2000 simulated samples in which the MLEs based on PIR data are within the parameter space, with $K = 40$, $d = 0$, and ζ

scaled in terms of the interval length. Because we use numerical maximization, the results of the maximization routine are never precisely on the borders of the parameter space. We therefore use boundaries of $|\text{logit } \hat{\phi}_{PIR}| < 8$ and $|\log \hat{\zeta}_{PIR}| < 8$ to define the edges of the parameter space. The proportion of estimates falling within the parameter space is highest for moderate values of prevalence and incidence ($0.2 \leq \phi \leq 0.5$ and $0.1 \leq \zeta \leq 0.25$), decreases as prevalence becomes more extreme, and decreases as incidence becomes either less frequent (less than once per ten intervals) or more frequent (more than once per four intervals). Unlike the MTS estimators, the pattern of boundary estimates is asymmetric because PIR tends to reach ceiling levels when prevalence is large.

In addition to returning estimates that are on the edges of the parameter space, the MLEs based on PIR data have the further disadvantage of being somewhat sensitive to initialization values. The likelihood surface becomes very flat when the PIR scores are

Table 2

Proportion of 2000 simulated PIR samples ($K = 40$) in which $|\text{logit}(\hat{\phi}_{PIR})| < 8$ and $|\log(\hat{\zeta}_{PIR})| < 8$.

	$\zeta = 0.02$	0.05	0.1	0.2	0.25	0.4	0.5
$\phi = 0.1$	0.59	0.83	0.94	0.94	0.90	0.78	0.72
0.2	0.67	0.91	0.98	0.99	0.98	0.92	0.84
0.3	0.74	0.94	0.99	1.00	1.00	0.94	0.87
0.4	0.78	0.96	1.00	1.00	0.99	0.93	0.86
0.5	0.77	0.96	0.99	0.99	0.98	0.89	0.80
0.6	0.75	0.94	0.97	0.96	0.92	0.79	0.67
0.7	0.68	0.87	0.92	0.85	0.78	0.56	0.38
0.8	0.60	0.76	0.77	0.56	0.42	0.16	0.09
0.9	0.47	0.48	0.32	0.08	0.04	0.00	0.00

near ceiling or floor levels, making it difficult to numerically identify the maximum. For the implementation of the estimators in the accompanying R package, the consequence is that drastically different estimates can be returned depending on the initialization values of the algorithm. Together with the possibility of obtaining estimates on the edges of the parameter space, the numerical instability of the MLEs motivates our investigation of alternative estimators that incorporate penalty functions.

Whole Interval Recording

Consider observing a behavior stream generated by the Alternating Poisson Process and recording observations using whole interval recording. As with PIR, suppose that one observes K intervals, where each interval includes c seconds of active observation time followed by d seconds of recording time. Let W_k indicate the WIR score from interval k , corresponding to the time from t_k to $t_k + c$. Following the WIR system, $W_k = 1$ if the behavior occurs for the duration of the active portion of interval, and $W_k = 0$ otherwise. Formally,

$$W_k = I \left[c = \int_0^c Z(t_k + s) ds \right] \quad (9)$$

for $k = 1, \dots, K$.

Using the WIR system to score a state behavior is logically equivalent to using PIR to score the absence of the behavior. WIR data can therefore be modeled just as PIR data, after an appropriate change of parameters. Specifically, the log-likelihood for WIR data under the Alternating Poisson Process can be written in terms of the log-likelihood for PIR data as

$$l_{WIR}(\phi, \zeta | W_1 = w_1, \dots, W_K = w_K) = l_{PIR}(1 - \phi, \zeta | U_1 = 1 - w_1, \dots, U_K = 1 - w_K). \quad (10)$$

The equivalence of the two system implies that estimates of prevalence and incidence based on WIR data can be obtained using the algorithms developed for PIR.

Penalized likelihood estimators

The previous section has illustrated that maximum likelihood estimators derived from MTS data or from interval recording data have undesirable operating characteristics when based on a moderate sample size of $K = 40$ intervals. In this section, we consider the use of penalized likelihood estimators (PLEs) to stabilize the behavior of the estimators and improve their performance in moderately sized samples. PLEs are derived by maximizing the sum of the log likelihood and a penalty term that depends on the parameters. Penalized likelihood estimation has been applied to an array of statistical problems where maximum likelihood methods break down, such as estimation of logistic regression coefficients in small samples (Galindo-Garre, Vermunt, & Bergsma, 2004; Gelman, Jakulin, Pittau, & Su, 2008), estimation of correlation structure in high-dimensional multivariate regression models (Warton, 2008), and estimation of variance components in multi-level models when the number of highest-level units is small (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013).

PLEs can be derived from a Bayesian perspective, by assigning a prior distribution to the model parameters and taking the posterior mode (given the prior and the data) as a point estimator (Chung et al., 2013). The penalty function is thus equivalent to the log of the prior distribution. An advantage of PLEs is that they provide a convenient and coherent way to incorporate into the estimation process prior information about the behavior.

In most research contexts, we expect that prior knowledge about characteristics of the behavior will be more readily expressed in terms of the average event duration (μ) and average interim time (λ). We therefore consider a class of priors in which μ and λ follow independent gamma distributions, with

$$\mu \sim \text{Gamma}(\alpha_\mu, c\theta_\mu), \quad \lambda \sim \text{Gamma}(\alpha_\lambda, c\theta_\lambda),$$

with hyperparameters $\alpha_\mu, \alpha_\lambda > 1$ and $\theta_\mu, \theta_\lambda > 0$. Note that the priors on μ and λ are scaled in terms of the active interval length, so that they do not depend on the time unit

in which the parameters and interval length are measured. This class of priors has the useful property that the implied priors on prevalence (ϕ) and incidence (ζ) have familiar distributional forms. Specifically, when $\theta_\mu = \theta_\lambda = \theta$, it follows that

$$\phi \sim \text{Beta}(\alpha_\mu, \alpha_\lambda), \quad \zeta \sim \text{Gamma}^{-1}\left(\alpha_\mu + \alpha_\lambda, \frac{1}{c\theta}\right)$$

and that ϕ is independent of ζ .

The penalty function implied by these priors depends on how they are parameterized (i.e., based on priors for μ, λ or priors for ϕ, ζ). We recommend using the (μ, λ) parameterization because it reduces to zero when $\alpha_\mu = \alpha_\lambda = 1$ and $\theta_\mu = \theta_\lambda = \infty$, making the PLEs equivalent to the MLEs. With this parameterization, the penalty function has the form

$$p(\phi, \zeta) = (\alpha_\mu - 1) \log(\phi) + (\alpha_\lambda - 1) \log(1 - \phi) - (\alpha_\mu + \alpha_\lambda - 2) \log(\zeta) - \frac{\frac{\phi}{\theta_\mu} + \frac{1-\phi}{\theta_\lambda}}{c\zeta}. \quad (11)$$

Given data based on recording system $s \in \{MTS, PIR, WIR\}$, the PLEs $\tilde{\phi}_s, \tilde{\zeta}_s$ are defined as the values that maximize $l_s(\phi, \zeta) + p(\phi, \zeta)$.

Application of penalized likelihood estimators requires the analyst to choose values for the hyperparameters of the prior distribution. If one does not have specific prior knowledge regarding the characteristics of the behavior stream, it is prudent to choose hyperparameters that have little influence on the values of the PLEs. These weak priors imply a penalty function that is relatively flat, so that PLEs will correspond closely with the MLEs except when the data contain little information about the parameters. We suggest that $\alpha_\mu = \alpha_\lambda = 1.5$ and $\theta_\mu = \theta_\lambda = 10$ are reasonable default choices for hyperparameters. The priors are highest at $\mu = \lambda = 5c$, or an average event duration and an average interim time of 5 intervals; the inter-quartile ranges are from $6.1c$ to $20.5c$. The implied prior for prevalence is symmetric about $\phi = 0.5$; its use amounts to adding the information from observing one independent moment where the probability of observing behavior is 0.5. We examine the empirical performance of the PLEs with these default hyperparameters in the following subsections.

In some research contexts, one may have fairly strong prior knowledge about certain behavioral characteristics, which can be used to inform the choice of hyperparameters. For example, prior experience with a class of behavior may suggest that it is unlikely for the average event duration to be more than two intervals in length, and that the most likely value for the average event duration is $\frac{1}{2}$ an interval length. This suggests choosing $\alpha_\mu = 3$ and $\theta_\mu = \frac{1}{4}$, so that the prior mode of μ is $(\alpha_\mu - 1)c\theta_\mu = \frac{c}{2}$ and $\Pr(\mu > 2c) < 0.02$. Absent strong prior information about λ , one might use the default hyperparameters suggested above, taking $\alpha_\lambda = 1.5, \theta_\lambda = 10$.

Finite-sample performance

We noted in a previous section that maximum likelihood estimates derived from MTS or PIR data are not always well-defined, even when the number of intervals is moderate. To remedy this problem, we have proposed the use of penalized likelihood estimates that are always well-defined and numerically stable. These approaches to estimation should produce equivalent results when based on very long observation sessions with many intervals of data, but they may differ when the number of intervals is more limited.

Both maximum likelihood and penalized likelihood estimation represent alternatives to the standard method of summarizing intermittent behavioral observation data, which is to use the summary proportion of intervals with behavior. For MTS data, the summary proportion is an unbiased estimate of prevalence under a very broad class of data-generating models (Rogosa & Ghandour, 1991); however, the summary proportion may be less efficient than the PLE or MLE for prevalence under the Alternating Poisson Process model. It is therefore important to compare the accuracy of all of these estimators. For PIR data, the summary proportion is biased as an estimate of prevalence (Rogosa & Ghandour, 1991), while likelihood-based methods provide approximately unbiased estimates for sufficiently long observation sessions. However, for a fixed sample size, it remains to be seen whether likelihood-based estimates of

Table 3

Simulation design

Parameter	Definition	Levels	Min.	Step	Max.
ϕ	Prevalence	19	0.05	0.05	0.95
ζ	Incidence	10	0.05	0.05	0.50
K	Number of intervals	15	10	10	150
s	Recording system	2	MTS, PIR		

prevalence are more accurate than the PIR summary proportion.

In order to understand the operating characteristics of the MLEs and PLEs in samples with a finite number of intervals, we conducted a computer simulation study. The simulations examined two specific questions: First, for a given type of data (MTS or PIR), how does the accuracy of the PLEs compare to the accuracy of the MLEs (for both prevalence and incidence) and the accuracy of the summary proportion (for prevalence only)? Second, how large a sample is needed to obtain approximately unbiased estimates of prevalence or incidence using penalized likelihood methods?

Table 3 summarizes the simulation design. For all simulation conditions, we set $c = 1$ and $d = 0$, so that the values of incidence are scaled in terms of the length of an interval; for example, $\zeta = 0.1$ corresponds to an incidence of one new behavioral episode per ten intervals. The true prevalence of the behavior stream was varied across nearly its entire possible range. Incidence was varied between $\zeta = .05$ (one new behavior per 20 intervals) and $\zeta = .50$ (one new behavior every 2 intervals) because this represents a range of behaviors where intermittent recording procedures might feasibly be applied; in particular, PIR measurements would quickly approach ceiling levels when behaviors occur more frequently than once per two intervals. The number of intervals in the observation session was varied from 10 to 150. This range spans a variety of situations in which intermittent behavioral observation recording might be used, from a quick

observation in a classroom where the observer needs to capture the behaviors of several children to an intensive observation of a single child over the course of an entire class period. Finally, the simulated recording system was either MTS or PIR; WIR was omitted because it is equivalent to using PIR for the absence of the behavior.

The simulations were carried out using the `ARPObservation` package (Pustejovsky, 2014) for the R statistical computing environment (R Core Team, 2014). For each combination of parameter values (ϕ, ζ, K) and recording system, we generated 10000 behavior streams of length K from an Alternating Poisson Process, then simulated data based on the recording system. For each string of simulated data, we calculated the summary proportion of intervals and then found the MLEs and PLEs using numerical maximization. For the PLEs, we used the default priors of $\alpha_\mu = \alpha_\lambda = 1.5$ and $\theta_\mu = \theta_\lambda = 10$.

We examined the performance of the estimators for four different parameterizations. We were interested in both the natural parameterizations of ϕ and ζ , as well as the logit of ϕ , $\log(\phi/(1-\phi))$, and $\log(\zeta)$. In the natural parameterization of ϕ , the logit of ϕ , and the log of ζ we examined bias, $E[\hat{\theta} - \theta]$, and root mean squared error (RMSE), $\sqrt{Var(\hat{\theta}) + (Bias(\hat{\theta}))^2}$, where θ is the parameter of interest. In the natural parameterization of ζ we examined the relative bias and the relative RMSE:

$$Bias_{rel}(\zeta) = \frac{E(\hat{\zeta} - \zeta)}{\zeta}, \quad RMSE_{rel}(\zeta) = \frac{\sqrt{Var(\hat{\zeta}) + (Bias(\hat{\zeta}))^2}}{\zeta}.$$

I feel like probably I have made an inadvertent mistake from lack of knowledge in here and left out a hat or failed to include one so please check carefully

When describing the estimators, have set varying criteria for what we believe constitutes “approximately unbiased.” In the case of ϕ this is an absolute bias less than 0.03 and in the case of $\text{logit}(\phi)$ an absolute bias less than 0.05. Lacking other estimators for incidence, we have set rather liberal criteria in the cases of ζ and $\log(\zeta)$. In the case of ζ it is a relative absolute bias of less than 0.10 and in the case of $\log(\zeta)$ an absolute bias of less than 0.10.

Why these criteria?
Is "approximately unbiased" actually what I should be saying here??

Results: MTS

Figure 1 illustrates the distribution of RMSE of the MTS MLEs, PLEs, and proportions for each parameterization across varying levels of K . Given that the MTS proportions are known to be unbiased estimates of prevalence, those estimates provide a point of reference for the accuracy of the MLEs and PLEs. Across all levels of K , the MLEs of ϕ have comparable RMSEs to the MTS proportions. However, the MLEs perform notably worse in terms of $\text{logit}(\phi)$ until $K = 120$ or more, a very large number of intervals. Across all levels of K , the PLEs of ϕ and $\text{logit}(\phi)$ have RMSEs comparable to or smaller than the MTS proportions. In terms of both ζ and $\text{log}(\zeta)$, the MLEs perform considerably worse across all levels of K than the PLEs and the corresponding estimates of prevalence. In contrast, the PLEs perform quite well for both ζ and $\text{log}(\zeta)$, with accuracy comparable to the accuracy of the corresponding estimates of prevalence.

Figure 2 illustrates the minimum value of K needed to meet the criteria we have set for “approximately unbiased” when utilizing the penalized estimators. In terms of ϕ and $\text{logit}(\phi)$, the estimator requires at most 40 intervals when $\zeta/c \geq 0.10$ across the entire range of ϕ . Even when ζ/c drops to 0.05, unbiased estimates of ϕ can be obtained across the entire range if $K = 100$ and unbiased estimates of $\text{logit}(\phi)$ can be obtained when $K = 60$. In terms of ζ , unbiased estimates can be obtained when K is at least 60 for $0.30 \leq \phi \leq 0.70$ and $\zeta \geq 0.10$. Unbiased estimates can also be obtained when K is at least 70 for $0.05 \leq \phi \leq 0.95$ and $0.15 \leq \zeta \leq 0.25$. In terms of $\text{log}(\zeta)$, unbiased estimates can be obtained when K is at least 70 and $0.40 \leq \phi \leq 0.60$ and $\zeta \geq 0.15$. Unbiased estimates can also be obtained when K is at least 60 and $0.15 \leq \phi \leq 0.85$ and $0.10 \leq \zeta \leq 0.25$.

Results: PIR

Figure 3 illustrates the distribution of RMSE of the PIR MLEs, PLEs, and proportions for each parameterization across varying levels of K . In contrast to MTS, we expect that the proportion estimates of prevalence for PIR will perform quite poorly.

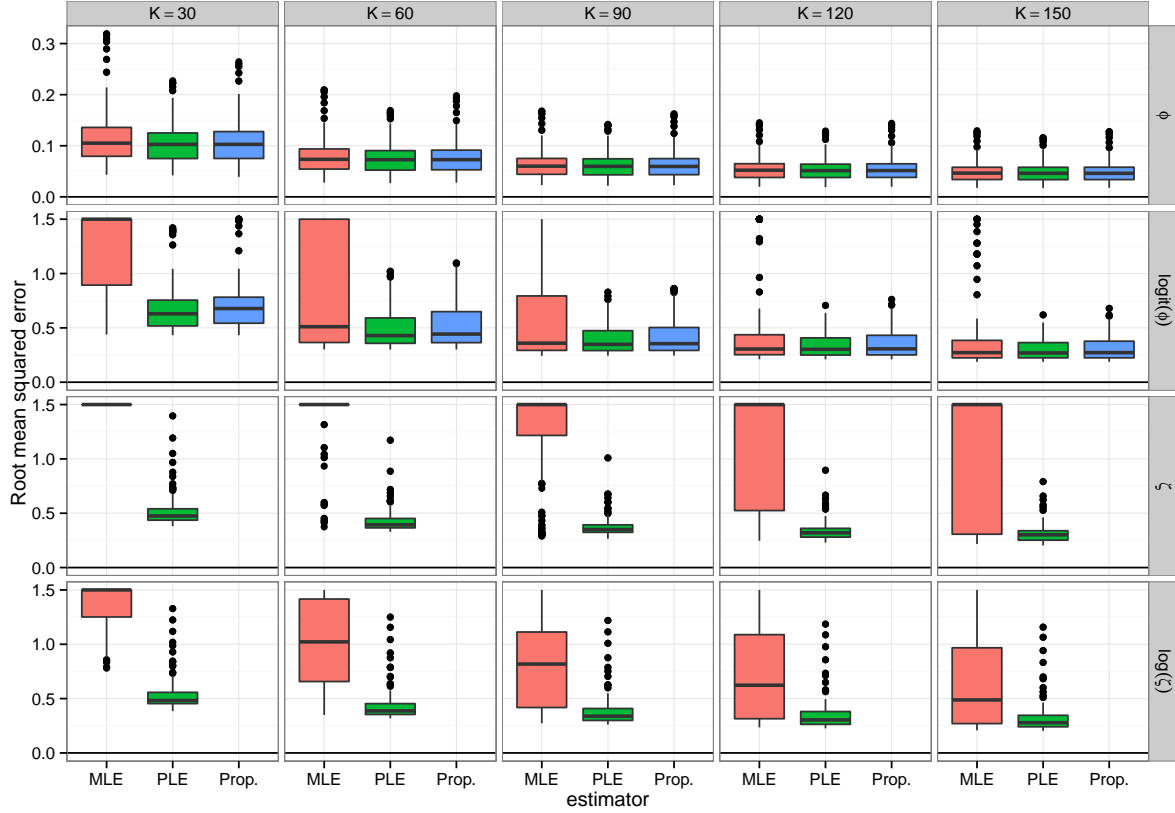


Figure 1. Distribution of root mean-squared error of MLE, PLE, and summary proportion (Prop.) estimates based on MTS data, across range of $0.05 \leq \phi \leq 0.95$ and $0.05 \leq \zeta \leq 0.50$. Values above 1.5 are not displayed. For zeta, the root mean-squared error is relative to the true value.

In fact, this is the case. In terms of the natural parameterization of ϕ , the MLE performs notably better than the proportions when $K \geq 60$, while the PLE performs better across the entire range of K . In terms of $\text{logit}(\phi)$, both the MLE and the proportion perform worse than the PLE. In both parameterizations of ϕ , the PLE performs just slightly worse for the PIR estimates than it did for the MTS estimates. In terms of the natural parameterization of ζ , the PLEs once again provide more accurate estimates of incidence. However, unlike with MTS data, the relative RMSE of the MLEs are not at ceiling even if the PLEs are clearly preferable. In terms of $\log(\zeta)$, the PLEs are definitely preferable and the RMSE of the MLEs once again range all the way up to

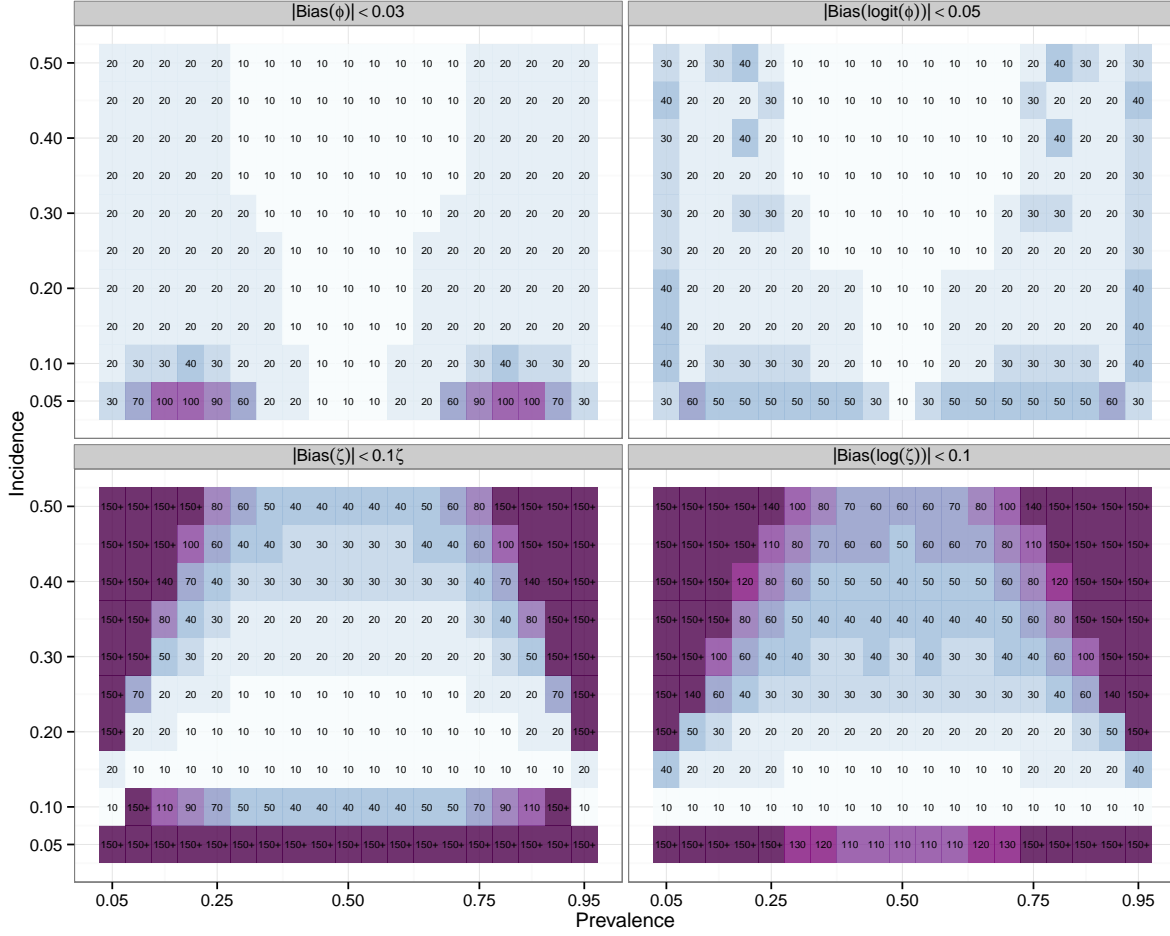


Figure 2. Minimum number of MTS intervals required to obtain PLEs with absolute bias less than given criteria.

the ceiling of 1.5 The mean RMSE of the PIR PLEs is similar to the MTS estimates of incidence for both parameterizations, but the spread when using PIR is larger.

Figure 4 illustrates the minimum value of K needed to meet the criteria we have set for “approximately unbiased” when utilizing the penalized estimators. The portion of the parameter space where one can obtain approximately unbiased estimates of prevalence with a moderate number of intervals is much smaller than with the MTS estimates. In terms of the bias of ϕ , unbiased estimates can be obtained when K is at least 70 and $0.10 \leq \phi \leq 0.35$ and $\zeta/c \leq 0.4$. Unbiased estimates can also be obtained when K is at least 60 and $\phi \leq 0.80$ and $\zeta/c \leq 0.25$. In terms of $\text{logit}(\phi)$, unbiased

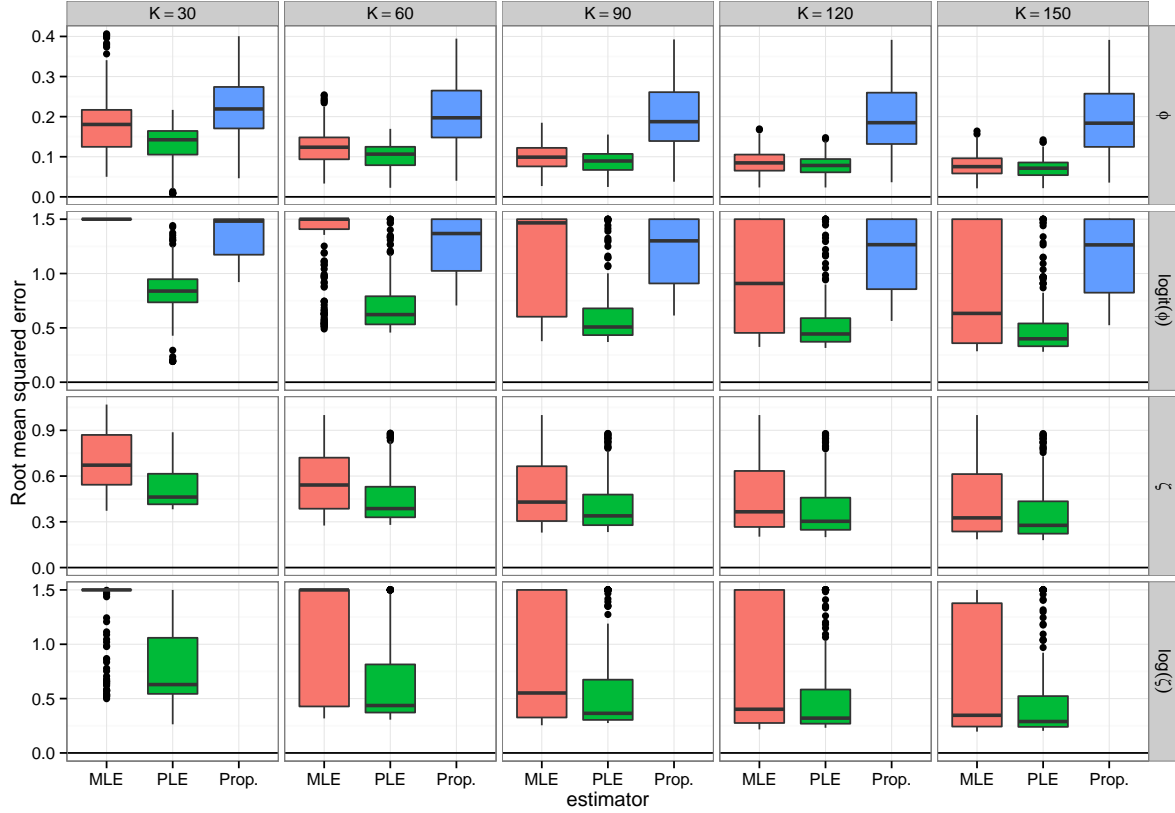


Figure 3. Distribution of root mean-squared error of MLE, PLE, and summary proportion (Prop.) estimates based on PIR data, across range of $0.05 \leq \phi \leq 0.95$ and $0.05 \leq \zeta \leq 0.50$. Values above 1.5 are not displayed. For zeta, the root mean-squared error is relative to the true value.

estimates can be obtained when K is at least 70 and $0.15 \leq \phi \leq 0.35$ and $\zeta/c \leq 0.30$. Unbiased estimates can also be obtained when K is at least 90 and $\phi \leq 0.70$ and $\zeta/c \leq 0.15$. In order to obtain unbiased estimates of incidence, much larger values of K are required. In terms of ζ , unbiased estimates can be obtained when K is at least 130 and $0.10 \leq \phi \leq 0.35$ and $\zeta \leq 0.45$. Unbiased estimates can also be obtained when K is at least 130, and $\phi \leq 0.60$ and $\zeta/c \leq 0.25$. In terms of $\log(\zeta)$, unbiased estimates can be obtained when K is at least 140 and $\phi \leq 0.40$ and $\zeta \leq 0.35$.

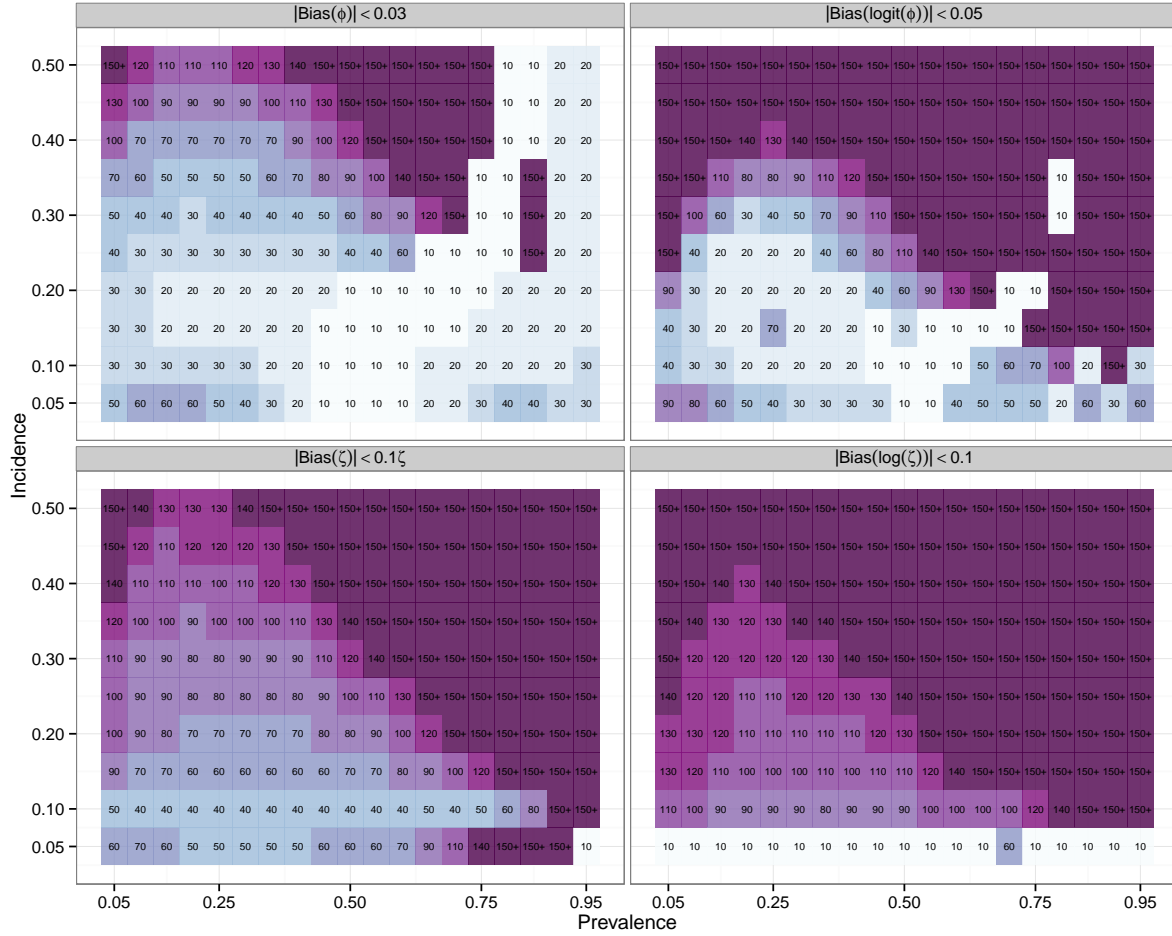


Figure 4. Minimum number of PIR intervals required to obtain PLEs with absolute bias less than given criteria.

Application

This section demonstrates the use of the PLEs with empirical behavioral observation data. The data are drawn from ?. In this study, trained raters watched six scripted, pre-recorded, ten-minute videos of a classroom during a lesson. The script outlined when students in the classroom were to act academically engaged or to act disruptively. For each video, academic engagement behavior was coded using three different recording systems: MTS, PIR, and WIR. For each of the three recording procedures, 15 s intervals were used, yielding a total of 40 intervals per scored clip. Each video clip was independently coded by ten different raters using each of the recording

systems.

The goal of our analysis is to estimate the prevalence and incidence of the behavior observed in each video clip, using the data collected using each of the three recording systems. Because each video was scored separately using each recording system, comparing the PLEs obtained from each type of data allows us to characterize the strengths and weaknesses of each recording system. The fact that each clip is scored by multiple raters also allows us to observe a source of measurement error that the Alternating Poisson Process model does capture: "inter-rater" error, due the human observers not perfectly following the scoring procedure or not perceiving the behavior stream with perfect accuracy.

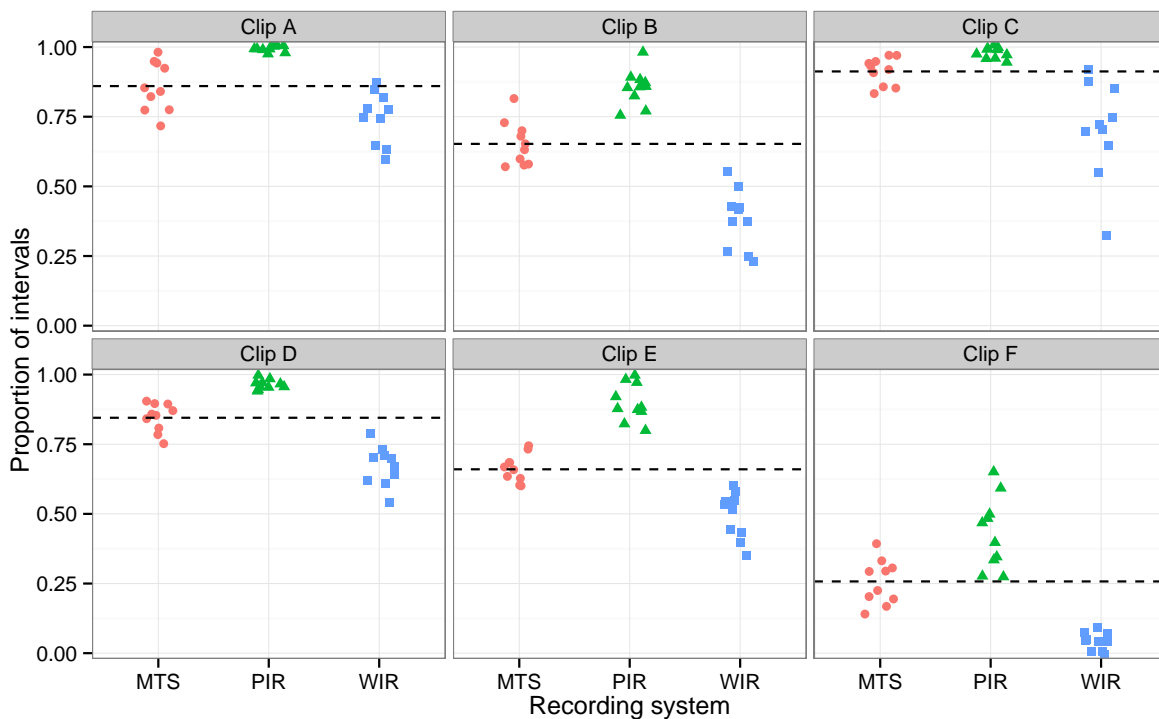


Figure 5. MTS, PIR, and WIR summary proportions for each rater, by clip. Horizontal dashed lines correspond to the average MTS summary proportions.

Figure 5 plots the overall proportion of intervals with academic engagment for each clip and each recording procedure, with separate points for each rater. The horizontal

dashed lines represent the average of the MTS proportions across all ten raters, which we treat as a benchmark for the prevalence estimates based on the Alternating Poisson Process. The PIR proportions are almost all higher than the unbiased MTS proportions, while the WIR proportions are nearly always lower, which is consistent with the fact that the PIR summary proportion is an upwardly biased estimate of prevalence while the WIR proportion is a downwardly biased estimate of prevalence. Note that the benchmark prevalence of academic engagement is high—above .80—in clips A, C, and D. In these three clips, the PIR proportions are frequently at or very near to the ceiling level. The raw proportions contain a suprisingly large amount of inter-rater variability.

Figure 6 displays the PLEs for prevalence based on the MTS, PIR, and WIR scores for each rater and each clip. The prevalence estimates are plotted on the vertical axis, with the corresponding summary proportion of intervals on the horizontal axis. The quality of the estimates varies substantially by both recording procedure and by clip. For MTS, the maximum likelihood estimates are all close to the raw proportions (which are unbiased estimates of prevalence). The MTS confidence intervals cover the benchmark prevalence estimate in 58 out of the 60 MTS records and are generally narrower than the confidence intervals based on the PIR and WIR data.

In all but clip F, the PIR estimates are highly unstable and essentially uninformative. We suspect that this is because PIR is a poor choice for recording the behavior in these clips, as indicated by the fact that the PIR proportions are at or near ceiling in these clips. In comparison, the WIR estimates for clips A, B, C, D, and E are relatively informative (though the CIs are still quite wide in some instances, particularly in clip B) This is consistent with theory that WIR is a more appropriate method of observing the behaviors in these clips, where academic engagement occurs for the majority of the observation session. Note that most of the confidence intervals cover the benchmark estimates of prevalence, and the average of the maximum likelihood estimates appears to be quite close to the benchmark.

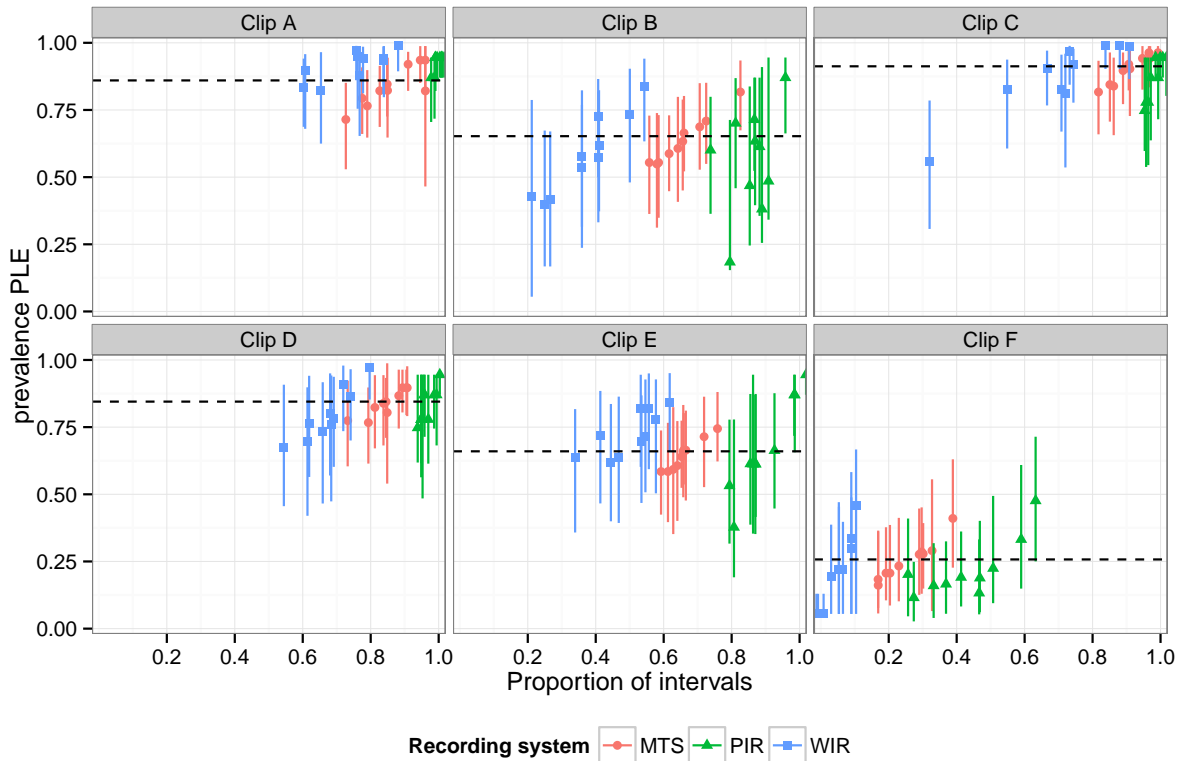


Figure 6. PLE estimates of prevalence versus proportion of intervals with behavior, for each recording system and each rater. Vertical whisker bars correspond to 95% confidence intervals for prevalence, obtained using a parametric bootstrapping procedure. Horizontal dashed lines correspond to the average MTS summary proportions.

For clip F, the situation is reversed: the PIR estimates are relatively informative while the WIR estimates are not, due to the relatively low level of academic engagement in this clip. Note that nearly all of the PIR confidence intervals cover the benchmark, and the average of the PIR point estimates is close to the benchmark estimate of prevalence.

A novel recording procedure

Thus far, we have considered conventional and widely used procedures for intermittent, systematic direct observation procedures. In this section, we describe a

recording procedure that can be used to obtain more accurate estimates of both prevalence and incidence. The method, which we call augmented interval recording, involves using MTS, PIR, and WIR systems on each interval. To the best of our knowledge, this procedure has not been previously described in the literature on systematic direct observation of behavior.

As in previous sections, suppose that the observation session is divided into K intervals and that the first c seconds of the interval are devoted to observation while the remaining d seconds are devoted to recording or resting; interval k therefore begins at time $t_k = (k - 1)(c + d)$.

Augmented interval recording

Consider an observer who uses the combination of MTS, PIR, and WIR scoring rules for each interval during an observation session. Doing so requires that the observer record sufficient data so that the values of the MTS, PIR, and WIR variables (X_{k-1}, U_k, W_k) for each interval. Figure 7 depicts the sequence of questions to be answered during interval k in order to completely determine these values. For each interval, the observer begins by noting the presence or absence of the behavior at time t_k and recording the MTS score. If the behavior is present ($X_{k-1} = 1$), then the partial interval record is also determined ($U_k = 1$), and it only remains to determine whether the behavior occurs for the duration of the interval ($W_k = 1$) or ends before the start of the next interval ($W_k = 0$). Similarly, if the behavior is absent at the start of the interval ($X_{k-1} = 0$), then the whole interval record is also determined ($W_k = 0$), and it only remains to determine whether a behavioral event begins before the start of the next interval ($U_k = 1$) or is absent for the entire interval ($U_k = 0$).

The AIR procedure requires only marginally more effort on the part of the observer than an interval recording method used alone. One measure of effort is the level of sustained attention required on the part of the observer. Because the sustained attention needed for interval recording also entails the attention needed for momentary

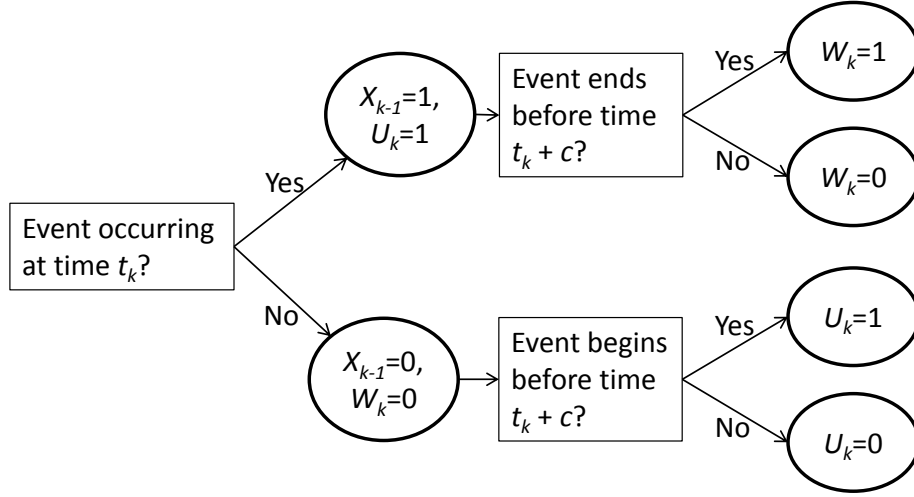


Figure 7. Procedure for combining MTS and interval recording

time sampling, the additional effort is minimal in this respect. Another measure of effort is the amount of data that must be recorded during the observation period. Because $W_k = 0$ is implied when $X_{k-1} = 0$ and $U_k = 1$ is implied when $X_{k-1} = 1$, AIR requires twice as much data as one of the single methods (rather than three times as much, might be supposed). Thus, for a fixed interval length, simultaneous use of all three methods entails at most twice as much effort as interval recording alone. Furthermore, using longer time-intervals with fewer intervals per observation period would mitigate the effort required.

Likelihood of AIR data

Under the assumptions of the Alternating Poisson Process, the data generated by the AIR system can be modeled using a discrete-time Markov Chain, from which estimates of prevalence and incidence can be obtained. The Markov property of the Alternating Poisson Process implies that the joint distribution can be written as

$$\begin{aligned} \Pr(X_0 = x_0, X_1 = x_1, U_1 = u_1, W_1 = w_1, \dots, X_K = x_K, U_K = u_K, W_K = w_K) \\ = \Pr(X_0 = x_0) \prod_{k=1}^K \Pr(X_k = x_k, U_k = u_k, W_k = w_k | X_{k-1} = x_{k-1}). \end{aligned} \quad (12)$$

Denote the transition probabilities $\Pr(X_k = b, U_k = c, W_k = d | X_{k-1} = a) = \pi_{a|bcd}$ and let $m_{a|bcd} = \sum_{k=1}^K I(X_{k-1} = a, X_k = b, U_k = c, W_k = d)$ for $a, b, c, d = 0, 1$. Conditional on X_0 , the log-likelihood of the observed AIR data is given by

$$l_{AIR}(\phi, \zeta) = \sum_{a=0}^1 \sum_{b=0}^1 \sum_{c=0}^1 \sum_{d=0}^1 m_{a|bcd} \log \pi_{a|bcd} \quad (13)$$

where

$$\begin{aligned} \pi_{0|000} &= [1 - p_0(d)] \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{0|010} &= 1 - p_0(c + d) - [1 - p_0(d)] \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{0|100} &= p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{0|110} &= p_0(c + d) - p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{1|010} &= 1 - p_1(c + d) - [1 - p_1(d)] \exp\left(\frac{-\zeta c}{\phi}\right) \\ \pi_{1|011} &= [1 - p_1(d)] \exp\left(\frac{-\zeta c}{\phi}\right) \\ \pi_{1|110} &= p_1(c + d) - p_1(d) \exp\left(\frac{-\zeta c}{\phi}\right) \\ \pi_{1|111} &= p_1(d) \exp\left(\frac{-\zeta c}{\phi}\right) \end{aligned}$$

and the remaining transition probabilities are all equal to zero. See Appendix B for the derivation of these quantities.

Relative efficiency

Discussion

Empirical evidence regarding assumptions of APP.

Robustness to violation of the APP assumptions.

Measurement model, needs to be extended to regression modeling, time series, random effects settings.

Use model to develop guidance about choice of recording procedures, interval length, rest time, K, etc.

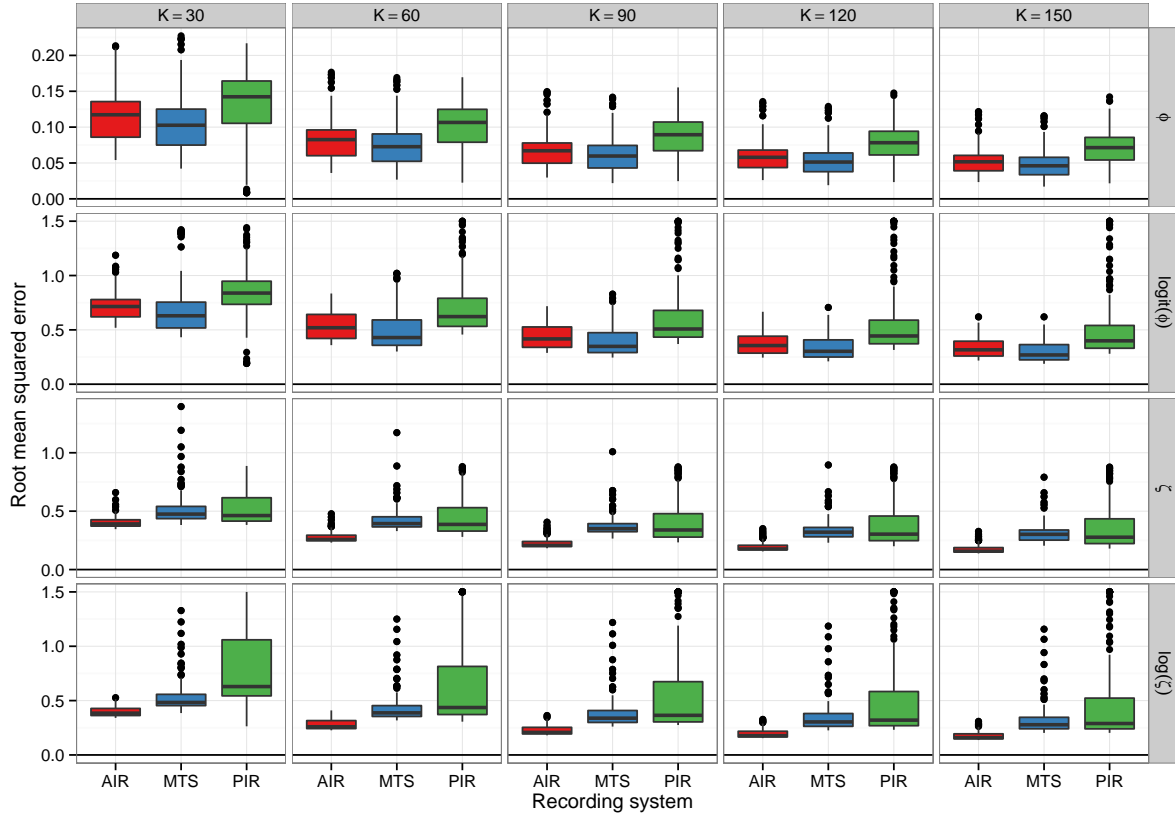


Figure 8. Distribution of root mean-squared error of PLE estimates based on AIR, MTS, and PIR data, across range of $0.05 \leq \phi \leq 0.95$ and $0.05 \leq \zeta \leq 0.50$. Values above 1.5 are not displayed. For zeta, the root mean-squared error for is relative to the true value.

This model does not account for observer error.

References

- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3/4), 227–267.
- Altmann, S. A., & Wagner, S. S. (1970). Estimating rates of behavior from Hansen frequencies. *Primates*, *11*(2), 181–183. doi: 10.1007/BF01731143
- Ayres, K., & Gast, D. L. (2010). Dependent measures and measurement procedures. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 129–165). New York, NY: Routledge.
- Bakeman, R., & Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. New York, NY: Cambridge University Press.
- Bornstein, M. H. (2002). Measurement variability in infant and maternal behavioral assessment. *Infant Behavior and Development*, *25*(4), 413–432. doi: 10.1016/S0163-6383(02)00143-1
- Brown, M., Solomon, H., & Stephens, M. A. (1977). Estimation of Parameters of Zero-One Processes by Interval Sampling. *Operations Research*, *25*(3), 493–505.
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. doi: 10.1007/s11336-013-9328-2
- Durand, V. M., Hieneman, M., Clarke, S., Wang, M., & Rinaldi, M. L. (2012). Positive Family Intervention for Severe Challenging Behavior I: A Multisite Randomized Clinical Trial. *Journal of Positive Behavior Interventions*, *15*(3), 133–143. doi: 10.1177/1098300712458324
- Galindo-Garre, F., Vermunt, J. K., & Bergsma, W. P. (2004, August). Bayesian Posterior Estimation of Logit Parameters with Small Samples. *Sociological Methods & Research*, *33*(1), 88–117. doi: 10.1177/0049124104265997
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of*

- Applied Statistics*, 2(4), 1360–1383. doi: 10.1214/08-AOAS191
- Griffin, B., & Adams, R. (1983). A parametric model for estimating prevalence, incidence, and mean bout duration from point sampling. *American Journal of Primatology*, 4(3), 261–271. doi: 10.1002/ajp.1350040305
- Haccou, P., & Meelis, E. (1992). *Statistical Analysis of Behavioural Data*. New York, NY: Oxford University Press.
- Hartmann, D. P., & Wood, D. D. (1990). Observational methods. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy* (2nd ed., pp. 107–138). New York, NY: Plenum Press.
- Kazdin, A. E. (2011). *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York, NY: Oxford University Press.
- Kraemer, H. C. (1979). One-zero sampling in the study of primate behavior. *Primates*, 20(2), 237–244.
- Kulkarni, V. G. (2010). *Modeling and Analysis of Stochastic Systems*. Boca Raton, FL: Chapman & Hall/CRC.
- Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early childhood special education and early intervention. *Topics in Early Childhood Special Education*. doi: 10.1177/0271121414524063
- Mann, J., Ten Have, T. R., Plunkett, J. W., & Meisels, S. J. (1991). Time sampling: A methodological critique. *Child Development*, 62(2), 227–241.
- Mudford, O. C., Taylor, S. A., & Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *Journal of Applied Behavior Analysis*, 42(1), 165–169. doi: 10.1901/jaba.2009.42-165
- Primavera, L. H., Allison, D. B., & Alfonso, V. C. (1996). Measurement of dependent variables. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 41–89). Mahwah, NJ: Lawrence Erlbaum.

- Pustejovsky, J. E. (2014). *ARPObservation: Simulating recording procedures for direct observation of behavior*. Retrieved from <http://cran.r-project.org/web/packages/ARPObservation>
- Pustejovsky, J. E., & Swan, D. M. (2014). *Four methods for analyzing partial interval recording data, with application to single-case research*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Rapp, J. T., Colby-Dirksen, A. M., Vollmer, T. R., Roane, H. S., Lomas, J., Britton, L. N., & Colby, A. M. (2007). Interval recording for duration events: A re-evaluation. *Behavioral Interventions*, 22, 319–345.
- Rogosa, D., & Ghandour, G. (1991). Statistical models for behavioral observations. *Journal of Educational Statistics*, 16(3), 157–252.
- Suen, H. K., & Ary, D. (1986, March). A post hoc correction procedure for systematic errors in time-sampling duration estimates. *Journal of Psychopathology and Behavioral Assessment*, 8(1), 31–38. doi: 10.1007/BF00960870
- Suen, H. K., & Ary, D. (1989). *Analyzing Quantitative Behavioral Observation Data*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Volpe, R., DiPerna, J., Hintze, J., & Shapiro, E. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, 34(4), 454–474.
- Warton, D. I. (2008). Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices. *Journal of the American Statistical Association*, 103(481), 340–349. doi: 10.1198/016214508000000021

Appendix A

Derivation of PIR model

The joint distribution of PIR observations depends on the conditional probabilities $\psi_k = \Pr [Z(t_k) = 1 | U_1, \dots, U_{k-1}]$. This appendix provides a derivation of Expression (7) in terms of the parameters of the Alternating Poisson Process. The derivation will make use of the following lemma.

Lemma 1. *The conditional probabilities of $Z(t_k) = 1, U_{k-1} = 1$ given $Z(t_{k-1})$ are:*

$$\begin{aligned} \Pr (Z(t_k) = 1, U_{k-1} = 1 | Z(t_{k-1}) = 1) &= p_1(c + d) \\ \Pr (Z(t_k) = 1, U_{k-1} = 1 | Z(t_{k-1}) = 0) &= p_0(c + d) - p_0(d) \exp \left(\frac{-\zeta c}{1 - \phi} \right). \end{aligned}$$

Proof. Observe that

$$\Pr (Z(t_k) = 1, U_{k-1} = 1 | Z(t_{k-1}) = 1) = \Pr (Z(t_k) = 1 | Z(t_{k-1}) = 1) = p_1(c + d)$$

and

$$\begin{aligned} \Pr (Z(t_k) = 1, U_{k-1} = 1 | Z(t_{k-1}) = 0) &= \int_0^c \frac{p_1(c-t)\zeta}{(1-\phi)} \exp \left(\frac{-\zeta t}{1-\phi} \right) dt \\ &= \phi \left[1 - \exp \left(\frac{-\zeta(c+d)}{\phi(1-\phi)} \right) - \exp \left(\frac{-\zeta c}{1-\phi} \right) + \exp \left(\frac{-\zeta(\phi c + d)}{\phi(1-\phi)} \right) \right] \\ &= p_0(c + d) - p_0(d) \exp \left(\frac{-\zeta c}{1-\phi} \right). \end{aligned}$$

□

Turning to the derivation of ψ_k , begin by noting that $U_{k-1} = 0$ implies that $Z(t_k + c) = 0$. It follows from the Markov property that

$$\begin{aligned} \Pr (Z(t_k) = 1 | U_1 = u_1, \dots, U_{k-2} = u_{k-2}, U_{k-1} = 0) \\ = \Pr (Z(t_k) = 1 | Z(t_{k-1} + c) = 0) = p_0(d). \end{aligned}$$

Next, Lemma 1 implies that

$$\begin{aligned} \Pr(Z(t_k) = 1, U_{k-1} = 1 | U_1, \dots, U_{k-2}) \\ = \psi_{k-1} p_1(c + d) + (1 - \psi_{k-1}) \left[p_0(c + d) - p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right) \right]. \end{aligned}$$

It therefore follows that

$$\begin{aligned} \Pr(Z(t_k) = 1 | U_1 = u_1, \dots, U_{k-2} = u_{k-2}, U_{k-1} = 1) \\ = \frac{\psi_{k-1} p_1(c + d) + (1 - \psi_{k-1}) \left[p_0(c + d) - p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right) \right]}{1 - (1 - \psi_{k-1}) \exp\left(\frac{-\zeta c}{1 - \phi}\right)}. \end{aligned}$$

Thus, ψ_k can be written as a function of ψ_{k-1} and u_{k-1} , as given in (7).

Appendix B

Derivation of AIR model

This appendix provides a derivation of the transition probabilities for the AIR model in terms of the parameters of the Alternating Poisson Process. Begin by noting that, by the definitions of the recording procedures, $X_{k-1} = 0$ implies that $W_k = 0$ and $X_{k-1} = 1$ implies that $U_k = 1$. It follows that $\pi_{0|bc1} = 0$ for $b, c = 0, 1$ and $\pi_{1|b0d} = 0$ for $b, d = 0, 1$. Derivation of the other transition probabilities will make use of the following lemma.

(The proof follow the same logic as in Lemma 1, and is therefore omitted.)

Lemma 2. *The conditional probability of $Z(t_k) = 1, W_{k-1} = 0$ given that $Z(t_{k-1}) = 1$ is*

$$\Pr(Z(t_k) = 1, W_{k-1} = 0 | Z(t_{k-1}) = 1) = p_1(c + d) - p_1(d) \exp\left(\frac{-\zeta c}{\phi}\right).$$

Turning to the eight remaining transition probabilities, note that

$$\begin{aligned} \pi_{0|100} &= \Pr(X_k = 1, U_k = 0 | X_{k-1} = 0) \\ &= \Pr(X_k = 1 | Z(t_k + c) = 0) \Pr(U_k = 0 | X_{k-1} = 0) \\ &= p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} \pi_{0|000} &= \Pr(X_k = 0, U_k = 0 | X_{k-1} = 0) = [1 - p_0(d)] \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{1|111} &= \Pr(X_k = 1, W_k = 1 | X_{k-1} = 1) = p_1(d) \exp\left(\frac{-\zeta c}{\phi}\right) \\ \pi_{1|011} &= \Pr(X_k = 0, W_k = 1 | X_{k-1} = 1) = [1 - p_1(d)] \exp\left(\frac{-\zeta c}{\phi}\right). \end{aligned}$$

Next, it follows from Lemmas 1 and 2 that

$$\begin{aligned} \pi_{0|110} &= \Pr(X_k = 1, U_k = 1 | X_{k-1} = 0) = p_0(c + d) - p_0(d) \exp\left(\frac{-\zeta c}{1 - \phi}\right) \\ \pi_{1|110} &= \Pr(X_k = 1, W_k = 0 | X_{k-1} = 1) = p_1(c + d) - p_1(d) \exp\left(\frac{-\zeta c}{\phi}\right). \end{aligned}$$

The two remaining transition probabilities can be obtained by subtraction:

$$\pi_{0|010} = 1 - \pi_{0|000} - \pi_{0|100} - \pi_{0|110} = 1 - p_0(c + d) - [1 - p_0(d)] \exp\left(\frac{-\zeta c}{1 - \phi}\right)$$

$$\pi_{1|010} = 1 - \pi_{1|011} - \pi_{1|110} - \pi_{1|111} = 1 - p_1(c + d) - [1 - p_1(d)] \exp\left(\frac{-\zeta c}{\phi}\right).$$