

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](mailto: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 code or record data from direct observation, including continuous recording, momentary time sampling (MTS), partial interval recording (PIR, also known as one-zero sampling, modified frequency sampling, Hansen sampling, or simply time-sampling), 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 a 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 measure changes in behavior over time and assess individual responses to intervention (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 (the frequency with which new episodes of the behavior begin, per time unit) and prevalence, or the proportion of time that the behavior occurs. Given an operationally defined behavior, measurements of its characteristics are obtained by observing the behavior (either in person, or by video-recording) for a specified length of time. The pattern of behavioral episodes over the course of an observation session—that is, what a human observer actually perceives—is often called the *behavior stream*.

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 event begins and ends. Data from continuous recording is very rich, in that it permits direct estimation of prevalence, incidence, and other aspects of the behavior stream; it 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), 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, the circumstances under which it is reasonable to reduce the data to a summary proportion depends on the recording system 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 may be quite reasonable if the investigator's interest is solely in the prevalence of the behavior. However, simple summaries of MTS data do not provide a clear measure of the behavior's incidence. Even if not of substantive interest, estimates of incidence are necessary for assessing the magnitude of measurement error in the prevalence estimates.

Brown, Solomon, and Stephens (1977) described a method for estimating both prevalence and incidence from MTS data. Their approach was to first posit a stochastic process for the underlying stream of behavior, as perceived by the observer, and then to consider the implications of observing the process intermittently via an MTS system. 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. The Alternating Poisson Process implies that interval-by-interval MTS scores follow a discrete-time Markov chain. Under the model, Brown et al. (1977, see also Griffin & Adams, 1983) provided closed-form expressions for the maximum likelihood estimators of prevalence and incidence and the associated asymptotic covariance matrix.

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 the incidence 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 the new behavioral episodes follow a Poisson process. While their 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

estimators 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 each episode of behavior has non-negligible duration. With state behaviors, the perceived behavior stream 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  $\mu$  to the sum of  $\mu$  and  $\lambda$  and the incidence of the behavior is equal to the reciprocal of the sum of  $\mu$  and  $\lambda$ . We will denote prevalence by  $\phi$ , where  $0 < \phi < 1$ , and incidence by  $\zeta$ , 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

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

Let  $\hat{p}_0 = n_{01} / (n_{00} + n_{01})$  and  $\hat{p}_1 = n_{11} / (n_{10} + n_{11})$ . Brown et al. (1977) showed that the maximum likelihood estimate of  $\zeta$  exists only when  $\hat{p}_0 < \hat{p}_1$ . When this condition holds, the maximum likelihood estimates of  $\phi$  and  $\zeta$  are given by

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

If  $\hat{p}_0 \geq \hat{p}_1$ , then the maximum likelihood estimator of  $\phi$  is still well-defined and equal to  $\hat{\phi}_{MTS} = (n_{01} + n_{11}) / K$ .

The probability that the maximum likelihood estimators are undefined or fall outside of the parameter space is not trivial, even when  $K$  is relatively large. Note that in order for the estimates to fall 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  $\zeta$  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 estimators are undefined is a substantial 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. For ease of notation, set  $\psi_1 = \phi$ , which follows from the assumption that the

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

process is in equilibrium. We show in Appendix A that

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

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  $\phi$  and  $\zeta$ , given observed PIR data  $u_1, \dots, u_K$ , is

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

Maximum likelihood estimators  $\hat{\phi}_{PIR}, \hat{\zeta}_{PIR}$  are obtained by maximizing  $l_{PIR}$  using a numerical algorithm. Because the conditional probabilities  $\psi_1, \dots, \psi_K$  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, with the likelihood expressed in terms of the unbounded parameters  $\text{logit}(\phi)$  and  $\log(\zeta)$ .

Just as with MTS data, maximum likelihood estimates 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  $\zeta$  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 maximum likelihood estimators for PIR data have the further disadvantage of being

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

somewhat sensitive to initialization values. The likelihood surface becomes very flat when the PIR scores are 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 algorithmic instability of the maximum likelihood estimators 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 maximum likelihood and penalized

likelihood estimators 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 also be derived from a Bayesian perspective, by assigning a prior distribution to the model parameters and taking their posterior mode (given the prior and the data) as a point estimator (Chung et al., 2013). In this approach, the penalty function is equivalent to the log of the prior distribution. In some contexts, the researcher may have prior knowledge about one or more characteristics of the behavior stream; for instance, it may be known that the average event duration is unlikely to be more than 15 s. In this situation, PLEs provide a convenient and coherent way to incorporate such prior information into the estimation process.

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 ( $\mu$ ) and average interim time ( $\lambda$ ). We therefore consider a class of priors in which  $\mu$  and  $\lambda$

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  $\mu$  and  $\lambda$  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 ( $\phi$ ) and incidence ( $\zeta$ ) 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  $\phi$  is independent of  $\zeta$ .

The penalty function implied by these priors depends on how they are parameterized (i.e., based on priors for  $\mu, \lambda$  or priors for  $\phi, \zeta$ ). We recommend using the  $\mu - \lambda$  parameterization because it reduces to zero when  $\alpha_\mu = \alpha_\lambda = 1$  and  $\theta_\mu = \theta_\lambda = \infty$ , making the PLEs equivalent to the maximum likelihood estimators. 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 then defined as the maximum of  $l_s(\phi, \zeta) + p(\phi, \zeta)$ .

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 maximum likelihood estimators 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 two independent moments, one with behavior and one without behavior. 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  $\mu$  is  $(\alpha_\mu - 1)c\theta_\mu = \frac{c}{2}$  and  $\Pr(\mu > 2c) < 0.02$ . Absent strong prior information about  $\lambda$ , one might use the default hyperparameters suggested above, taking  $\alpha_\lambda = 1.5, \theta_\lambda = 10$ .

#### **Finite-sample performance: MTS**

#### **Finite-sample performance: PIR**

### **Application**

#### **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 1 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 ( $U_k = 0$ ).

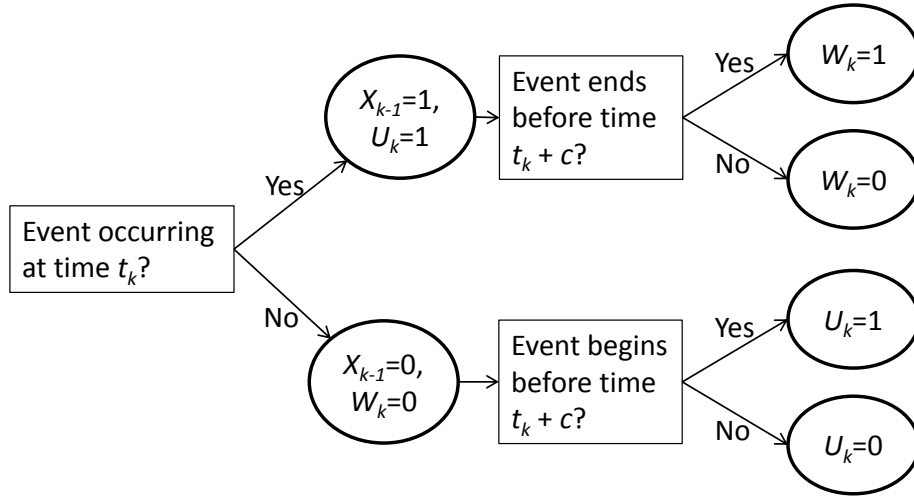


Figure 1. Procedure for combining MTS and interval recording

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 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.

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., & 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.
- 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  $\psi_k$ , 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,  $\psi_k$  can be written as a function of  $\psi_{k-1}$  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:

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

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