Old and novel methods for recording direct observations of a state behavior: Models and

efficiency considerations

James E. Pustejovsky

The University of Texas at Austin

Author Note

James E. Pustejovsky, 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 in many areas of educational and psychological research, particularly in applied research areas such as the treatment of behavioral disorders. A number of different procuedres are used to record data during direct observation, including continuous recording, momentary time sampling (MTS), partial interval recording (PIR), and whole interval recording (WIR). Methodological research on the relative advantages of different recording procedures is complicated by the fact that their properties depend on both the prevalence and the incidence of the behavior under observation. Furthermore, as typically summarized, PIR and WIR data measure neither of these characteristics. This paper proposes an Alternating Poisson Process model for interval recording data, which permits estimation of both prevalence and incidence by maximum likelihood. The paper also describes two novel observation recording methods that involve combinations of MTS, PIR, and WIR and that provides considerably more efficient estimates of incidence. The asymptotic relative efficiency of different recording procedures is studied under the Alternating Poisson Process.

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

Old and novel methods for recording direct observations of a state behavior: Models and efficiency considerations

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 operationally 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 (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 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 and incidence and 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, three conventional and widely used 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.

This paper examines a model for PIR and WIR data, from which principled estimates of prevalence and incidence can be obtained. Following Brown, Solomon, and Stephens (1977), I model the underlying behavior stream using an Alternating Poisson Process and then derive a model for the interval-by-interval scores. I then describe two novel procedures for intermittent recording of a behavior, each of which entails combining MTS and interval recording methods, and derive corresponding models based on an Alternating Poisson Process. Finally, I compare the efficiency of prevalence and incidence estimates based on each recording procedure and provide guidance regarding the relative advantages of the old and novel methods.

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 I will call episode durations, and spans of time in between episodes, which I 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 episode is occurring at time t and $Z(t) = 0$ otherwise.

The Alternating Poisson Process makes several assumptions about how the behavior stream is generated. Specifically, it is assumed that episode durations and interim times are mutually independent, random quantities, that the episode 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. 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 can be parameterized in a number of equivalent ways, any of which might be useful under certain circumstances. For instance, the behavior could be described in terms of its mean episode duration and mean interim time (μ, λ), in terms of its prevalence and incidence (ϕ, ζ), or in terms of its prevalence and its total transition rate between states, $\rho = 1/\mu + 1/\lambda = \zeta/\phi(1 - \phi)$. For ease of notation, I use the latter parameterization in describing the models for different recording procedures, but use the prevalence-incidence parameterization to study the relative efficiency of those procedures. I discuss translations between parameterizations in a later section.

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 $0 \leq r < s$:

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

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.,

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

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:

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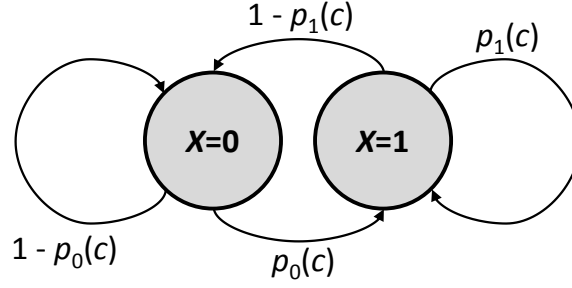\end{aligned} \quad (3)$$

Figure 1. Transition diagram for momentary time sampling model

(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 $X_0, X_1, ..., X_K$. 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 $k = 0, ..., K$.

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 $a = 0, 1$. Figure 1 depicts the transition diagram for this process. 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 $k = 1, ..., K$; 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

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

Brown et al. (1977) provided closed-form expressions for the maximum likelihood estimators (MLEs) for ϕ and ρ 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 ρ exists only when $\hat{p}_0 < \hat{p}_1$. When this condition holds, the MLEs for ϕ and ρ are given by

$$\hat{\phi}_M = \frac{\hat{p}_0}{\hat{p}_0 + 1 - \hat{p}_1} \qquad \text{and} \qquad \hat{\rho}_M = \frac{-\log(\hat{p}_1 - \hat{p}_0)}{c}. \tag{5}$$

Any other comments on MTS?

In a later section, the asymptotic efficiency of the MLEs based on MTS data will be compared to that of the corresponding MLEs based on other recording procedures. These comparisions involve the Fisher information, which is defined as the negative expectation of the Hessian of the log-likelihood. Appendix A provides further details regarding calculation of the Fisher information for the MTS model.

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 is the behavior occurs at any point during the active portion of interval, and $U_k = 0$ otherwise. In terms of the behavior stream process,

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

for $k = 1, ..., K$, 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 $U_1, ..., U_K$ can be derived as follows. Let $\psi_k, k = 2, ..., K$ denote the probability than 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. I show in Appendix B that

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

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

$$
\Pr\left(U_k = 1 \,|\, Z(t_k) = 0\right) = 1 - e^{-\rho\phi c}.
$$

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

$$
\Pr\left(U_k = 1 \,|\, U_1, ..., U_{k-1}\right) = \psi_k \Pr\left(U_k = 1 \,|\, Z(t_k) = 1)\right) + (1 - \psi_k)\Pr\left(U_k = 1 \,|\, Z(t_k) = 0)\right)
$$

$$
= 1 - (1 - \psi_k)e^{-\rho\phi c}.
$$

The joint distribution of $U_1, ..., U_K$ can therefore be expressed as

$$
\Pr\left(U_1 = u_1, ..., U_K = u_K\right) = \Pr\left(U_1 = u_1\right) \prod_{k=2}^{K} \Pr\left(U_k = u_k \,|\, U_1, ..., U_{k-1}\right)
$$

$$
= \prod_{k=1}^{K} \left[1 - (1-\psi_k)e^{-\rho\phi c}\right]^{u_k} \left[(1-\psi_k)e^{-\rho\phi c}\right]^{(1-u_k)}. \tag{8}
$$

The log-likelihood of ϕ and ζ, given observed PIR data $u_1, ..., u_K$, is thus

$$
l_{PIR}(\phi, \zeta) = \sum_{k=1}^{K} u_k \ln\left[1 - (1-\psi_k)e^{-\rho\phi c}\right] + (1-u_k)\left[\ln(1-\psi_k) - \rho\phi c\right]. \tag{9}
$$

Let $\hat{\phi}_{PIR}$ and $\hat{\zeta}_{PIR}$ denote the maximum likelihood estimators based on PIR data, defined as the values that maximize l_{PIR}. Unlike with MTS data, the estimators based on PIR data do not have closed-form expressions, but can be found using a numerical optimization algorithm.

Although the log-likelihood in Equation (9) is useful for computing the maximum likelihood estimators, it is otherwise rather difficult to analyze. A different representation of the data reveals the fact that the model is a discrete-time Markov Chain process, which facilitates computation of the Fisher information. Let the random
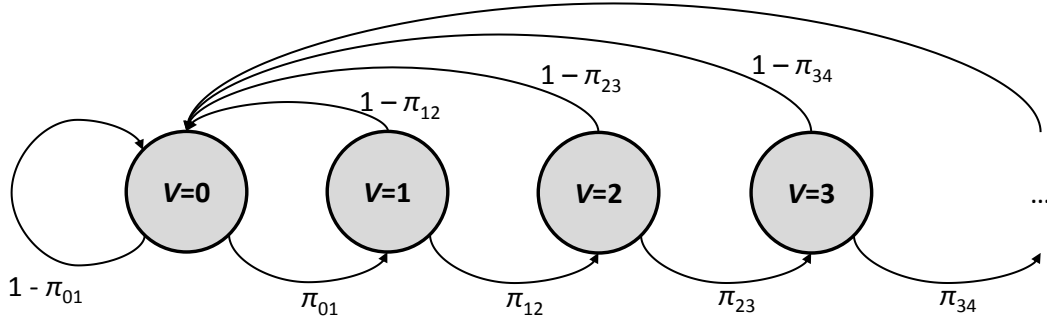
Figure 2. Transition diagram for partial interval recording model

variable V_0 represent the state of this process immediately prior to the start of the observation session, and define $V_k = (V_{k-1} + 1) U_k$ for $k = 1, ..., K$. (Note that the PIR interval data can be recovered from this representation by taking $U_k = I(V_k - V_{k-1} > 0)$.) Observe that if there is no behavior during interval $k - 1$, then ψ_k does not depend on the data from any previous intervals. More generally, ψ_k depends only on the number of preceding intervals in which a behavior occurred. It follows that

$$\Pr\left(V_k = a | V_0 = v_0, ..., V_{k-1} = v\right) = \Pr\left(V_k = a | V_{k-1} = v\right)$$

for $a \in \{0, v + 1\}$. Thus, $V_1, ..., V_K$ follow a discrete-time Markov Chain process on the non-negative integers. Figure 2 depicts the transition diagram for this process. The transition probabilities are given by

$$\Pr\left(V_k = v + 1 | V_{k-1} = v\right) = \pi_{v,v+1} = 1 - \left[1 - \overset{v}{\circ} f\left(p_0(d)\right)\right] e^{-\rho \phi c}$$

and $\Pr\left(V_k = 0 | V_{k-1} = v\right) = 1 - \pi_{v,v+1}$, where

$$f(q) = \frac{q p_1(c + d) + (1 - q)\left[p_0(c + d) - p_0(d)e^{-\rho \phi c}\right]}{1 - (1 - q)e^{-\rho \phi c}}$$

and $\overset{v}{\circ} f$ denotes the v-fold recursion of f.

Comment on expected information for PIR

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 is the behavior occurs for the duration of the active portion of interval, and $W_k = 0$ otherwise. Formally,

$$W_k = I \left[c = \int_0^c Z \left(t_k + s \right) ds \right] \tag{10}$$

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

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

$$l_{WIR} \left(\phi, \zeta | W_1 = w_1, ..., W_k = w_k \right) = l_{PIR} \left(1 - \phi, \zeta | U_1 = 1 - w_1, ..., U_K = 1 - w_k \right). \tag{11}$$

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.

Augmented interval recording

Thus far, I have considered three conventional and widely used procedures for intermittent, systematic direct observation procedures—MTS, PIR, and WIR. These three procedures are closely related, and the logical dependencies between the scoring rules suggest that it might be both feasible and useful to use them in combination. To the extent that each scoring rule captures different information about the behavior stream, then combining all three might procedure more efficient estimates of the behavioral parameters. Thus, I now consider a novel recording procedure, which I will call augmented interval recording, that involves applying all three scoring rules simultaneously to each interval in the observation session. To the best of my knowledge,

this procedures has not been previously described in the literature on systematic direct observation of behavior.

Just as with PIR or WIR, 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 $t_k = (k-1)(c+d)$.

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) can be determined for each interval. Figure 3 depicts the sequence of questions to be answered during interval k in order to completely determine these values. For each interval, the observer first notes the presence or absence of the behavior at time t_k and records 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 (in which case $W_k = 1$) or ends before time $t_k + c$ (in which case $W_k = 0$). 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 time $t_k + c$ (in which case $U_k = 1$) or is absent for the entire interval (in which case $U_k = 0$).

The AIR procedure might require 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, as
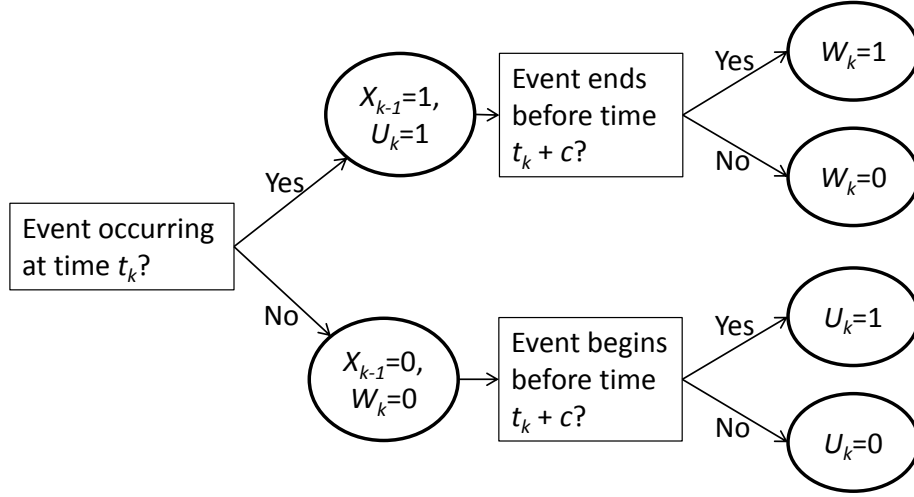
Figure 3. Procedure for combining MTS and interval recording

might be supposed). Thus, for a fixed interval length, simultaneous use of all three methods would entail 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.

Under the assumptions of the Alternating Poisson Process, the data generated by the AIR system can be modeled using a 6-state 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 of the measurements can be written as

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

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

$$l_{AIR}\left(\phi, \zeta\right) = \sum_{a,b,c,d} m_{a|bcd} \log \pi_{a|bcd} \quad (13)$$
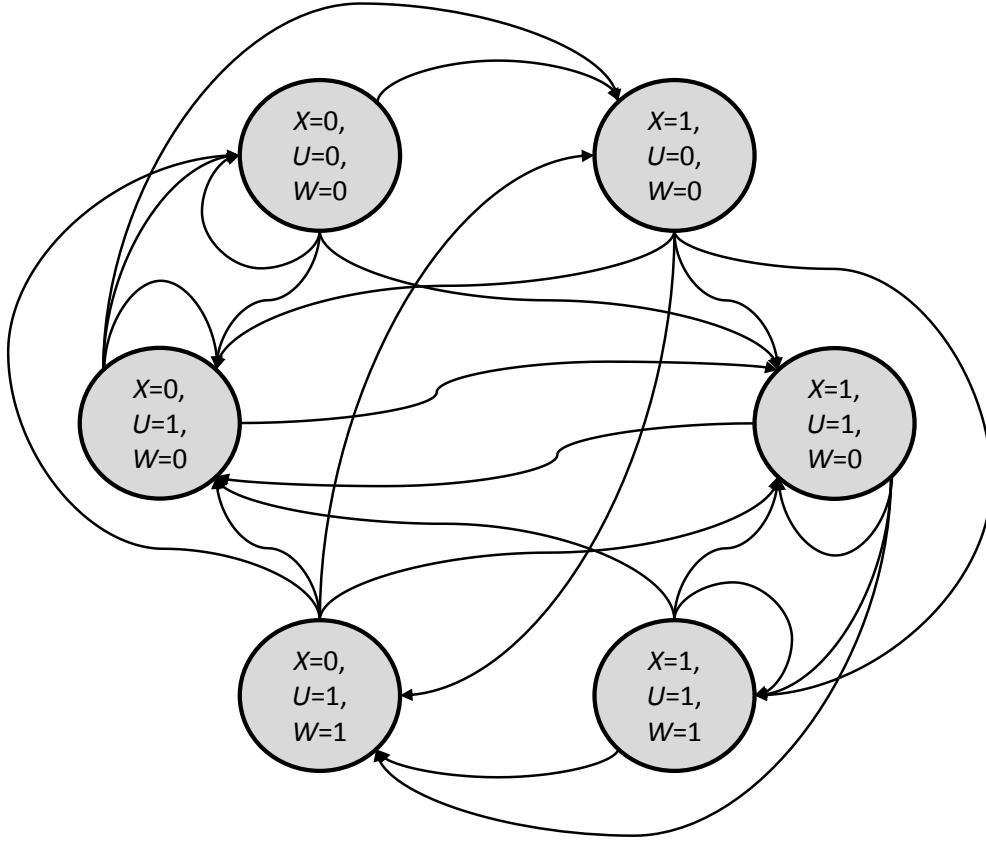
Figure 4. Transition diagram for augmented interval recording model. Transition probabilities are omitted for clarity.

where

$$\pi_{0|000} = [1 - p_0(d)] \, e^{-\rho\phi c}$$

$$\pi_{0|100} = p_0(d)e^{-\rho\phi c}$$

$$\pi_{0|010} = 1 - p_0(c+d) - [1 - p_0(d)] \, e^{-\rho\phi c}$$

$$\pi_{0|110} = p_0(c+d) - p_0(d)e^{-\rho\phi c}$$

$$\pi_{1|010} = 1 - p_1(c+d) - [1 - p_1(d)] \, e^{-\rho(1-\phi)c}$$

$$\pi_{1|110} = p_1(c+d) - p_1(d)e^{-\rho(1-\phi)c}$$

$$\pi_{1|011} = [1 - p_1(d)] \, e^{-\rho(1-\phi)c}$$

$$\pi_{1|111} = p_1(d)e^{-\rho(1-\phi)c},$$

the remaining transition probabilities are all equal to zero, and the sum is taken over the terms with strictly positive transition probabilities. See Appendix C for the derivation of these quantities. Figure 4 depicts the transition diagram for this process. As with PIR data, the MLEs $\hat{\phi}_{AIR}, \hat{\zeta}_{AIR}$ are obtained by maximizing l_{AIR} using numerical optimization techniques.

Comment on expected information for AIR

Intermittent transition recording

Discussion

In this paper, we have considered how to estimate the prevalence and incidence of a state behavior from data collected using conventional intermittent observation recording systems, including momentary time sampling, partial interval recording, and whole interval recording. Following earlier work by Brown et al. (1977) on MTS, we used an Alternating Poisson Process to model the behavior stream as perceived by the observer, from which models for PIR and WIR data can be derived. For estimating the model parameters, simulation evidence indicated that penalized likelihood methods with generic, weak priors generally outperformed maximum likelihood methods–often dramatically so. For PIR data, penalized likelihood estimates of prevalence provided much more accurate estimates than the naivë summary proportion, which is currently widely used.

We have also described a novel recording procedure, augmented interval recording, that involves combining MTS, PIR, and WIR. For a given period of observation, and using intervals twice the length of other procedures, AIR provides estimates of prevalence that are only slightly less accurate than estimates based on MTS, while also providing estimates of incidence that are substantially more accurate than estimates based on any other procedure. Of course, at present these advantages are only theoretical. To determine whether AIR offers any advantage in practice, its feasibility in real-life research contexts will need to be assessed.

Across all of the recording procedures that we have considered, the foremost
limitation of the proposed models and estimation techniques is the strength of the
assumptions entailed by the Alternating Poisson Process model for the behavior stream.
The model posits that the individual episodes of behavior and spans of time in between
episodes are exponentially distributed. Whether these distributional assumptions are
reasonable–and for what classes of behavior–is an important question requiring further
empirical research. Addressing it will likely require measuring the behavior of a large
sample of participants using intensive, continuous recording techniques. Another related
avenue of further research is to examine the extent to which the proposed estimation
techniques are robust to violations of the distributional assumptions (e.g., assuming that
event durations follow a gamma distribution that has lower variance than the
exponential). It may well be that the robustness of the PLEs depends on which system
is used to record the data, and whether prevalence or incidence is of primary interest.

Several other limitations of these models should also be acknowledged. Our
approach has treated the recording procedures themselves as essentially mechanical
procedures that can be applied without human error, yet in practice the procedures are
not perfectly reliable–as evidenced by the high level of inter-rater variability in the **?**
data. Arguably, the model we have considered could be interpretted as implicitly
accommodating rater error by allowing that the Alternating Poisson Process describes
the observer's *perception* of the behavior stream, rather than the true behavior stream.
However, a more explicit approach to accounting for human error in the recording
process would be more useful. How to extend the model in this respect remains an open
question for further research.

Another limitation of the models is that they are limited to describing
measurement error from a single observation session. In practice, systematic behavioral
observation data is often collected on a single participant across many sessions (as in a
single-case study) or across many participants (as in a between-subjects experiment). In

either setting, it would be useful to embed the measurement model that we have proposed in a generalized linear modeling framework, which could be used to describe changes in prevalence and incidence across time, or in a random-effects framework, which could be used to describe between-subjects variation in the characteristics of behavior streams.

Despite these limitations, the models and estimation methods that we have proposed are nonetheless useful. For MTS data, penalized likelihood provides a means to estimate incidence as well as to assess the extent of measurement error in the estimate of prevalence. For PIR data, the penalized likelihood estimates of prevalence represent an improvement over the current standard approach, which is to simply ignore the bias in the summary proportion. Though not the focus of the present paper, the models that we have described can also be applied to develop better psychometric guidance for behavioral observation data. Through further mathematical analysis or though computer simulation, the models that we have presented could be used to study how choices regarding recording procedures, interval lengths, rest times, and observation session lengths influence the precision of behavioral measurements. Guidance regarding these aspects of study design would be useful to applied researchers designing single-case experiments or between-subjects trials.

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3/4), 227–267.

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.

Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, *32*(1), 12–40.

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.

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

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.

Kulkarni, V. G. (2010). *Modeling and Analysis of Stochastic Systems*. Boca Raton, FL:

Chapman & Hall/CRC.

Mann, J., Ten Have, T. R., Plunkett, J. W., & Meisels, S. J. (1991). Time sampling: A methodological critique. *Child Development*, *62*(2), 227–241.

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.

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.

Appendix A

Fisher information matrix for MTS model

The Fisher information for the MTS model and other models can be calculated using some general results on structurally parameterized discrete-time Markov chains. Consider a stationary and ergodic Markov chain with finite state-space $\mathcal{S}$, transition probabilities π_{ab} for $a, b \in \mathcal{S}$, and equilibrium distribution $\tilde{\pi}_a$ for $a \in \mathcal{S}$. Suppose that the transition probabilities are functions of a vector-valued parameter $\boldsymbol{\theta}$. Under some regularity conditions, the Fisher information is then given by

$$\mathcal{I}_{\boldsymbol{\theta}} = K \sum_{a,b \in \mathcal{S}} \frac{\tilde{\pi}_a}{\pi_{ab}} \frac{\partial \pi_{ab}}{\partial \boldsymbol{\theta}} \frac{\partial \pi_{ab}}{\partial \boldsymbol{\theta}'}, \tag{14}$$

where the sum is taken over the transition probabilities that are strictly positive Billingsley (1961). Thus, the Fisher information for a given model can be calculated from the transition probabilities, derivatives of the transition probabilities, and equilibrium distribution.

> Does the same hold for countably infinite state spaces?

For the MTS model, the state space is $\{0, 1\}$, the equilibrium distribution is given by $\tilde{\pi}_0 = 1 - \phi, \tilde{\pi}_1 = \phi$, and the derivatives of the transition probabilities with respect to ϕ and ρ are

$$\frac{\partial \pi_{00}}{\partial \phi} = -\left[1 - \exp(-\rho c)\right] \qquad \frac{\partial \pi_{00}}{\partial \rho} = -\phi c \exp(-\rho c)$$

$$\frac{\partial \pi_{01}}{\partial \phi} = 1 - \exp(-\rho c) \qquad \frac{\partial \pi_{01}}{\partial \rho} = \phi c \exp(-\rho c)$$

$$\frac{\partial \pi_{10}}{\partial \phi} = -\left[1 - \exp(-\rho c)\right] \qquad \frac{\partial \pi_{10}}{\partial \rho} = (1 - \phi) c \exp(-\rho c)$$

$$\frac{\partial \pi_{11}}{\partial \phi} = 1 - \exp(-\rho c) \qquad \frac{\partial \pi_{11}}{\partial \rho} = -(1 - \phi) c \exp(-\rho c).$$

Appendix B

Derivation of PIR model

The joint distribution of PIR observations depends on the conditional probabilities

$\psi_k = \Pr\left[Z(t_k) = 1 \,|U_1, ..., U_{k-1}\right]$. 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:*

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

$$\Pr\left(Z(t_k) = 1, U_{k-1} = 1 \,|Z(t_{k-1}) = 0\right) = p_0(c + d) - p_0(d)e^{-\rho\phi c}.$$

Proof. Observe that

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

and

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

$\square$

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

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

Next, Lemma 1 implies that

$$\Pr\left(Z(t_k) = 1, U_{k-1} = 1 \,|\, U_1, ..., U_{k-2}\right)$$

$$= \psi_{k-1} p_1(c+d) + (1 - \psi_{k-1}) \left[p_0(c+d) - p_0(d) e^{-\rho \phi c} \right].$$

It therefore follows that

$$\Pr\left(Z(t_k) = 1 \,|\, U_1 = u_1, ..., U_{k-2} = u_{k-2}, U_{k-1} = 1\right)$$

$$= \frac{\psi_{k-1} p_1(c+d) + (1 - \psi_{k-1}) \left[p_0(c+d) - p_0(d) e^{-\rho \phi c} \right]}{1 - (1 - \psi_{k-1}) e^{-\rho \phi c}}.$$

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

Expected information matrix for PIR.

Appendix C

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*

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

Turning to the eight remaining transition probabilities, note that

$$\pi_{0|100} = \Pr\left(X_k = 1, U_k = 0 | X_{k-1} = 0\right)$$

$$= \Pr\left(X_k = 1 | Z(t_k + c) = 0\right) \Pr\left(U_k = 0 | X_{k-1} = 0\right)$$

$$= p_0(d)e^{-\rho\phi c}.$$

Similarly,

$$\pi_{0|000} = \Pr\left(X_k = 0, U_k = 0 | X_{k-1} = 0\right) = \left[1 - p_0(d)\right] e^{-\rho\phi c}$$

$$\pi_{1|111} = \Pr\left(X_k = 1, W_k = 1 | X_{k-1} = 1\right) = p_1(d)e^{-\rho(1-\phi)c}$$

$$\pi_{1|011} = \Pr\left(X_k = 0, W_k = 1 | X_{k-1} = 1\right) = \left[1 - p_1(d)\right] e^{-\rho(1-\phi)c}.$$

Next, it follows from Lemmas 1 and 2 that

$$\pi_{0|110} = \Pr\left(X_k = 1, U_k = 1 | X_{k-1} = 0\right) = p_0(c + d) - p_0(d)e^{-\rho\phi c}$$

$$\pi_{1|110} = \Pr\left(X_k = 1, W_k = 0 | X_{k-1} = 1\right) = p_1(c + d) - p_1(d)e^{-\rho(1-\phi)c}.$$

The two remaining transition probabilities can be obtained by subtraction:

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

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

The information matrix for the AIR model can be calculated from Equation (14), with equilibrium probabilities $\Pr(X_k = 0) = 1 - \phi$ and $\Pr(X_k = 1) = \phi$. The derivatives of the transition probabilities with respect to ϕ and ρ are as follows:

$$\frac{\partial \pi_{0|000}}{\partial \phi} = -\rho c \pi_{0|000} - e^{-\rho \phi c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{0|100}}{\partial \phi} = -\rho c \pi_{0|100} + e^{-\rho \phi c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{0|010}}{\partial \phi} = -\left(1 - e^{-\rho(c+d)}\right) + \rho c \pi_{0|000} + e^{-\rho \phi c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{0|110}}{\partial \phi} = \left(1 - e^{-\rho(c+d)}\right) + \rho c \pi_{0|100} - e^{-\rho \phi c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{1|010}}{\partial \phi} = -\left(1 - e^{-\rho(c+d)}\right) - \rho c \pi_{1|011} + e^{-\rho(1-\phi)c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{1|110}}{\partial \phi} = \left(1 - e^{-\rho(c+d)}\right) - \rho c \pi_{1|111} - e^{-\rho(1-\phi)c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{1|011}}{\partial \phi} = \rho c \pi_{1|011} - e^{-\rho(1-\phi)c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{1|111}}{\partial \phi} = \rho c \pi_{1|111} + e^{-\rho(1-\phi)c}\left(1 - e^{-\rho d}\right)$$

$$\frac{\partial \pi_{0|000}}{\partial \rho} = -\phi c \pi_{0|000} - \phi d e^{-\rho(\phi c + d)}$$

$$\frac{\partial \pi_{0|100}}{\partial \rho} = -\phi c \pi_{0|100} + \phi d e^{-\rho(\phi c + d)}$$

$$\frac{\partial \pi_{0|010}}{\partial \rho} = -\phi(c+d)e^{-\rho(c+d)} + \phi c \pi_{0|000} + \phi d e^{-\rho(\phi c + d)}$$

$$\frac{\partial \pi_{0|110}}{\partial \rho} = \phi(c+d)e^{-\rho(c+d)} + \phi c \pi_{0|000} - \phi d e^{-\rho(\phi c + d)}$$

$$\frac{\partial \pi_{1|010}}{\partial \rho} = (1-\phi)(c+d)e^{-\rho(c+d)} + (1-\phi)c\phi_{1|011} - (1-\phi)d e^{-\rho[(1-\phi)c + d]}$$

$$\frac{\partial \pi_{1|110}}{\partial \rho} = -(1-\phi)(c+d)e^{-\rho(c+d)} + (1-\phi)c\phi_{1|111} + (1-\phi)d e^{-\rho[(1-\phi)c + d]}$$

$$\frac{\partial \pi_{1|011}}{\partial \rho} = -(1-\phi)c\phi_{1|011} + (1-\phi)d e^{-\rho[(1-\phi)c + d]}$$

$$\frac{\partial \pi_{1|111}}{\partial \rho} = -(1-\phi)c\phi_{1|111} - (1-\phi)d e^{-\rho[(1-\phi)c + d]}$$