

Direct observation recording: Algorithms used in **ARPObservation**

James E. Pustejovsky

May 24, 2014

This vignette describes the algorithms used in **ARPObservation** to simulate behavior streams and direct observation recording data based on an alternating renewal process. The ARP is a statistical model that can be used to describe the characteristics of simple behavior streams, in which a behavior of interest is either occurring or not occurring at a given point in time. We will refer to the length of individual episodes of behavior as *event durations* and the lengths of time between episodes of behavior as *interim times*. In the ARP framework, variability is introduced into the behavior stream by treating each individual event duration and each interim time as a random quantity, drawn from some probability distribution.

The sequence of events comprising the behavior stream can be described as follows. Let L denote the length of the observation session. Let A_1 denote the duration of the first behavioral event observed, A_2 denote the duration of the second event, and A_u the duration of event u for $u = 3, 4, 5, \dots$. Let B_0 denote the length of time from the beginning of the observation session until the first behavioral event, with $B_0 = 0$ if event 1 is occurring at the beginning of the session. Let B_u denote the u^{th} interim time, meaning the length of time between the end of event u and the beginning of event $u+1$, for $u = 1, 2, 3, \dots$. The values $B_0, A_1, B_1, A_2, B_2, A_3, B_3, \dots$ provide a quantitative description of the behavior stream from an observation session. Note that these quantities are measured in time units, such as seconds.

1 Simulating the behavior stream

1.1 Generating distributions

The **ARPObservation** package generates behavior streams that follow an alternating renewal process with specified generating distributions. The package provides two approaches to generating the initial interim time and initial event duration, which we explain further below. Subsequent event durations A_2, A_3, A_4, \dots are generated independently, following a specified probability distribution with mean μ and cumulative distribution function (cdf) $F(t; \mu)$. Sub-

sequent interim times B_1, B_2, B_3, \dots are generated independently, following a specified probability distribution with mean λ and cdf $G(t; \lambda)$.

The package currently includes functions for exponential distributions, gamma distributions, mixtures of two gamma distributions, Weibull distributions, uniform distributions, and constant values. Each distribution is implemented as an object of class `eq-dist`, which provides functions for generating random deviates from the specified distribution and the corresponding equilibrium distribution. For distributions involving more than a single parameter, all parameters except for the mean must be specified.

1.2 Initial conditions

ARPObservation provides two approaches to generating the initial interim time and initial event duration. The first approach involves the following steps:

1. Generate a random number X from a Bernoulli distribution with user-specified probability p_0 .
2. If $X = 0$, then generate B_0 from the same distribution as subsequent interim times, i.e., from $G(t; \lambda)$. If $X = 1$, then set $B_0 = 0$.
3. Generate A_1 from the same distribution as subsequent event durations, i.e., from $F(t; \mu)$.

If p_0 is not specified by the user, the default value of $p_0 = 0$ is used, so that behavior streams always begin with an interim time. This approach produces behavior streams that are initially out of equilibrium.

The other approach uses initial conditions chosen so that the resulting process is in equilibrium. This involves the following steps:

1. Generate a random number X from a Bernoulli distribution with probability $\mu/(\mu + \lambda)$.
2. If $X = 0$, then generate B_0 from the distribution with cdf

$$\tilde{G}(t; \lambda) = \frac{1}{\lambda} \int_0^t [1 - G(x; \lambda)] dx.$$

If $X = 1$, then set $B_0 = 0$.

3. If $X = 0$, then generate A_1 from $F(t; \mu)$. If $X = 1$, then generate A_1 from the distribution with cdf

$$\tilde{F}(t; \mu) = \frac{1}{\mu} \int_0^t [1 - F(x; \mu)] dx.$$

2 Direct observation recording procedures

The package provides several algorithms that simulate commonly used direct observation recording procedures. Each algorithm takes as input a randomly generated behavior stream and produces as output a summary measurement (or measurements) from a direct observation procedure.

2.1 Event counting

Event counting produces a measurement Y^E equal to the number of events that begin during the observation session. Let J denote the number of last behavioral event that begins during the observation session, which can be calculated by finding the integer that satisfies the inequalities

$$\sum_{j=0}^{J-1} (A_j + B_j) \leq L < \sum_{j=0}^J (A_j + B_j),$$

where we define $A_0 = 0$ for notational convenience. It follows that $Y^E = J$.

2.2 Continuous recording

Continuous recording produces a measurement Y^C equal to the proportion of the observation session during which the behavior occurs. In order to calculate this quantity from the behavior stream, we must account for the possibility that the last event beginning during the observation session may have a duration that extends beyond when the session ends. The measurement based on continuous recording can be calculated as

$$Y^C = \begin{cases} \frac{1}{L} \sum_{j=1}^J A_j & \text{if } \sum_{j=1}^J (B_{j-1} + A_j) \leq L \\ 1 - \frac{1}{L} \sum_{j=0}^{J-1} B_j & \text{if } \sum_{j=1}^J (B_{j-1} + A_j) > L \end{cases}$$

2.3 Momentary time sampling

In momentary time sampling, an observer divides the observation session into K intervals of equal length and notes whether the behavior is present or absent at the very end of each interval. The summary measurement Y^M then corresponds to the proportion of moments during which the behavior is observed. Let $X_k = 1$ if the behavior is occurring at the end of interval k for $k = 1, \dots, K$. The value of X_k can be calculated from the behavior stream as follows. Let $I(X)$ denote the indicator function, equal to one if condition X is true and zero otherwise. Let m_k be the number of the last event that ends before the k^{th} interval ends, defined formally as

$$m_k = \sum_{i=1}^J I \left[\sum_{j=1}^i (B_{j-1} + A_j) < kL \right]$$

for $k = 1, \dots, K$. If interim time B_{m_k} concludes before the end of interval k (or equivalently, if event A_{m_k+1} begins before the end of interval k), then $X_k = 1$; formally,

$$X_k = I \left[\sum_{j=0}^{m_k} (A_j + B_j) < kL \right]$$

for $k = 1, \dots, K$. The summary measurement is then calculated as $Y^M = \sum_{k=1}^K X_k / K$.

2.4 Partial interval recording

Like momentary time sampling, partial interval recording is also based on a set of K intervals of equal length, but a different rule is used to score each interval. In partial interval recording, the observer counts a behavior as present if it occurs at any point during the first c time units of the interval, where $c \leq L/K$; the remaining $L/K - c$ time units are used to record notes or rest. Let $U_k = 1$ if the behavior occurs at any point during the k^{th} interval, $U_k = 0$ otherwise. The k^{th} interval will be equal to one if and only if interim time m_{k-1} ends during the active part of the interval. Noting that interim time m_{k-1} ends at time $\sum_{j=0}^{m_{k-1}} (A_j + B_j)$ and that the active part of the k^{th} interval ends at time $(k-1)L + c$, it can be seen that

$$U_k = I \left[\sum_{j=0}^{m_{k-1}} (A_j + B_j) < (k-1)L + c \right],$$

for $k = 1, \dots, K$. The summary measurement Y^P is then calculated as the proportion of intervals during which the behavior is observed at any point:

$$Y^P = \sum_{k=1}^K U_k / K.$$

2.5 Whole interval recording

Whole interval recording is similar to partial interval recording but uses yet a different rule to score each interval. Specifically, the observer counts a behavior as present only if it occurs for all c time units at the beginning of the interval. Let $W_k = 1$ if the behavior occurs for the duration, with $W_k = 0$ otherwise. Let n_k be the number of the last event that begins before the k^{th} interval begins, defined formally as

$$n_k = \sum_{i=1}^J I \left[\sum_{j=0}^i (A_j + B_j) < (k-1)L \right]$$

for $k = 1, \dots, K$. It follows that W_k will be equal to one if and only if event n_k ends after the active part of interval k :

$$W_k = I \left[\sum_{j=1}^{n_k} (B_{j-1} + A_j) \geq (k-1)L + c \right],$$

for $k = 1, \dots, K$. The summary measurement Y^W is then calculated as the proportion of intervals during which the behavior is observed at any point:

$$Y^W = \sum_{k=1}^K W_k / K.$$