To Jongmin and Nairi re observation in finite experimental time-window of processes of random onset and duration

Henrik Flyvbjerg

March 12, 2019

Contents

1 Formulation of problem **1**

2 Statistics of Class-I observations **2**

3 Statistics of Class-II observations **3**

3.1 General $p(\Delta t)$ 3

3.2 Exponential $p(\Delta t)$ 4

3.2.1 Estimators 4

3.2.2 Biases of estimators 4

3.2.3 Variances of estimators 4

4 Statistics of Class-III observations **5**

5 Statistics of Class-IV observations **5**

1 Formulation of problem

We observe individual binding and unbinding events. We record the duration, Δt, of every bound state we observe, duration being the time passing from binding occurred at time t_{b} till unbinding occurred at time t_{u}. We make these observations in an experimental time-window of fixed duration τ. Consequently, we cannot observe t_{b} and t_{u} for every bound state we do observe. Thus, every observed bound state belongs to one of four classes:

I. Bound states with binding occurring before we start observing, and unbinding occurring while we observe, $t_\mathrm{b} < 0 < t_\mathrm{u} < \tau$.

II. Bound states with binding occurring while we observe and unbinding occurring while we observe, $0 < t_\mathrm{b} < t_\mathrm{u} < \tau$.

III. Bound states with binding occurring while we observe and unbinding occurring after we stop observing, $0 < t_\mathrm{b} < \tau < t_\mathrm{u}$. ending while we observe.

IV. Bound states with binding occurring before we start observing and unbinding occurring after we stop observing, $t_\mathrm{b} < 0 < \tau < t_\mathrm{u}$.

We assume that individual binding events occur at random with a rate that is constant in time, both before and during during observation, i.e., binding is a Poisson process. We assume that once binding has occurred, unbinding is a random event that occurs after a duration Δt that does not depend on when binding occurred. Then the probability distribution of durations, $p(\Delta t) = p(t_\mathrm{u} - t_\mathrm{b})$, does not depend on t_b.

Normalization and notation:

$$\begin{aligned} \int_0^\infty p(s) ds &= 1 \ , && (1) \\ P(\Delta t) &\equiv \int_0^{\Delta t} p(s) ds \ , && (2) \\ Q(\Delta t) &\equiv \int_{\Delta t}^{\infty} p(s) ds \ , && (3) \\ P(\Delta t) + Q(\Delta t) &= 1 \ . && (4) \end{aligned}$$

For an exponential distribution,

$$p(\Delta t) = r e^{-r\Delta t} \ , \quad (5)$$

$$\begin{aligned} P(\Delta t) &= 1 - e^{-r\Delta t} \ , && (6) \\ Q(\Delta t) &= e^{-r\Delta t} \ . && (7) \end{aligned}$$

2 Statistics of Class-I observations

Class I observations accumulate statistics only for observations with of $t_\mathrm{b} < 0 < t_\mathrm{u} < \tau$. This results in an observed distribution of unbinding times, t_u, call it $p_\mathrm{I}(t_\mathrm{u})$. It relative values are

$$p_\mathrm{I}(t_\mathrm{u}) \propto Q(t_\mathrm{u})\theta(\tau - t_\mathrm{u}) \approx Q(t_\mathrm{u}) \ . \quad (8)$$

Here $\theta(t)$ is Heaviside's step-function: $\theta(t) = 1$ for $t > 0$ and $\theta(t) = 0$ for $t < 0$. Its value at $t = 0$ does not matter in the integrals it occurs in: Points have zero measure. The approximation relies on $p(\Delta t) \approx 0$ for $\Delta t > \tau$.

Normalized for Class-I observations only,

$$p_{\mathrm{I}}(t_{\mathrm{u}}) = \frac{Q(t_{\mathrm{u}})\theta(\tau - t_{\mathrm{u}})}{\int_0^\tau Q(s)ds} \approx \frac{Q(t_{\mathrm{u}})}{\int_0^\infty Q(s)ds} \quad , \tag{9}$$

If $p(\Delta t)$ is the exponential distribution in Eq. 5,

$$p_{\mathrm{I}}(t_{\mathrm{u}}) = \frac{re^{-rt_{\mathrm{u}}}}{1 - e^{-r\tau}} \approx re^{-rt_{\mathrm{u}}} \quad . \tag{10}$$

We see that r can be estimated from Class-I observations alone. Since r can be estimated also from other observations, as we shall see below, the optimal estimator will be one that combines information from all observations.

3 Statistics of Class-II observations

Class II observations accumulate statistics only for values of $\Delta t < \tau$ and that with a bias. Thus, the distribution $p_{\mathrm{II}}(\Delta t)$ of these observed durations differs from $p(\Delta t)$.

3.1 General $p(\Delta t)$

The bias arises because Class-II observations of duration Δt must start with $t_{\mathrm{b}} \in]0, \tau - \Delta t[$ for $t_{\mathrm{u}} < \tau$ to be satisfied. Since t_{b} has uniform probability density on the time axis and Δt does no depend on it,

$$p_{\mathrm{II}}(\Delta t) \propto (\tau - \Delta t)p(\Delta t)\theta(\tau - \Delta t) \quad . \tag{11}$$

Normalized for Class-II observations only,

$$p_{\mathrm{II}}(\Delta t) = \frac{(\tau - \Delta t)p(\Delta t)\theta(\tau - \Delta t)}{\tau P(\tau) - \int_0^\tau s\, p(s)ds} \quad . \tag{12}$$

To the extent that $p(\Delta t) \approx 0$ for $\Delta t > \tau$, $p(\Delta t)\theta(\tau - \Delta t) = p(\Delta t)$ and

$$p_{\mathrm{II}}(\Delta t) = \frac{\tau - \Delta t}{\tau - E[\Delta t]} p(\Delta t) = \frac{1 - \Delta t/\tau}{1 - E[\Delta t]/\tau} p(\Delta t) \quad , \tag{13}$$

where we have used $P(\tau) = 1$ and introduced the expected value $E[\cdot]$ of Δt.

3.2 Exponential $p(\Delta t)$

3.2.1 Estimators

For $p(\Delta t)$ the exponential distribution in Eq. 5, Eq. 12 reads

$$p_{\mathrm{II}}(\Delta t) = \frac{(\tau - \Delta t)\theta(\tau - \Delta t) r^2 e^{-r\Delta t}}{r\tau - 1 + e^{-r\tau}} \approx \frac{r^2(\tau - \Delta t)}{r\tau - 1} e^{-r\Delta t} \quad . \tag{14}$$

Maximum likelihood estimation of r based on N independent observations of Δt, all distributed according to $p_{\mathrm{II}}(\Delta t)$, gives the following equation for the estimate r:

$$r = (\overline{\Delta t})^{-1} \frac{r\tau - 2 + (r\tau + 2)e^{-r\tau}}{r\tau - 1 + e^{-r\tau}} \approx (\overline{\Delta t})^{-1} \frac{r\tau - 2}{r\tau - 1} \tag{15}$$

where $\overline{\Delta t}$ is the simple average of the observed durations.

Given $\overline{\Delta t}$ and τ, r may be found numerically from this transcendental equation, possibly by iterating it, starting with the approximation $r = 1/\overline{\Delta t}$.

In the last approximation, the equation for r is quadratic, so r can be found analytically as a function of $\overline{\Delta t}$ and τ:

$$r = \frac{1}{2\overline{\Delta t}} \left(1 + \overline{\Delta t}/\tau + \sqrt{1 - 6\overline{\Delta t}/\tau + \left(\overline{\Delta t}/\tau\right)^2} \right) \quad . \tag{16}$$

We note that the square root in this solution becomes imaginary unless τ is about six times $\overline{\Delta t}$ or longer, which is a strangely demanding requirement when we consider how small that makes $\exp(-r\tau)$. It does indicate, however, that the exponential terms in the transcendental equation matter in a significant manner even for such values of $r\tau$ that makes them small. This issue is maybe best investigated numerically, by comparing the exact, but numerical, solution for r with the analytical, but approximate solution.[1]

3.2.2 Biases of estimators

Analytical work in progress...

3.2.3 Variances of estimators

Analytical work in progress...

[1] This is a nice exercise for Jongmin, but before he spends time on it, he should check my calculations for simple mistakes.

4 Statistics of Class-III observations

Class III observations accumulate statistics only for observations with of $0 < t_\mathrm{b} < \tau < t_\mathrm{u} <$. This results in an observed distribution of unbinding times, t_b, call it $p_\mathrm{III}(t_\mathrm{b})$. Its relative values are

$$p_\mathrm{III}(t_\mathrm{b}) \propto Q(\tau - t_\mathrm{b})\theta(t_\mathrm{b}) \approx Q(\tau - t_\mathrm{u}) \ . \tag{17}$$

The approximation relies on $p(\Delta t) \approx 0$ for $\Delta t > \tau$.

Normalized for Class-III observations only,

$$p_\mathrm{III}(t_\mathrm{b}) = \frac{Q(\tau - t_\mathrm{b})\theta(t_\mathrm{b})}{\int_0^\tau Q(s)ds} \approx \frac{Q(\tau - t_\mathrm{b})}{\int_0^\infty Q(s)ds} \ , \tag{18}$$

If $p(\Delta t)$ is the exponential distribution in Eq. 5,

$$p_\mathrm{III}(t_\mathrm{b}) = \frac{re^{-r(\tau - t_\mathrm{b})}}{1 - e^{-r\tau}} \approx re^{-r(\tau - t_\mathrm{b})} \ . \tag{19}$$

We see that r can be estimated also from Class-III observations alone. As already mentioned, since r can be estimated also from other observations, the optimal estimator will be one that combines information from all observations.

5 Statistics of Class-IV observations

It is a bad or hard-pressed experiment that has Class-IV observations at all. If one is forced by circumstances to make such a hard-pressed experiment, one needs to make the most of all observations, including Class-IV observations.

The only information that these observations provide is their total count, i.e., the number of bound states observed that lasted the entire time of observation. This count contributes information only to a statistics that compares it to the number of counts of observations of the other three classes.

To make this comparison, we calculate the probabilities that, given that an observation occurs in the experimental window, it belongs to each of the four disjoint classes. [Work in progress...]

6 Estimator using all observations

Analytical work in progress...