Events occurring in time, the Poisson distribution, and Scampy

Chris Wild
Department of Statistics, University of Auckland
September 2019

(*https://www.stat.auckland.ac.nz/~wild/scampy/*)
is a ***dynamic-graphics web app*** for either experiencing (or showing to students): the stochastic nature of events occurring over time, the behaviour of events occurring in time viewed both from a waiting-time perspective and from an occurrences-per-time-interval perspective, and also for experiencing how the two relate. There is too much going on for a single-webpage environment so developments are unveiled in story-board fashion. You will have to spend some time reading the below and getting familiar with using the app before sending your students to use it for activities.

[Why "***Scampy***"? It is made up from the names of the people most involved in its design: Stephanie Budget, Chris Wild, Anna Fergusson, Maxine Pfannkuch, and Yu Han Soh (Soh Yu Han, who also did all of the programming as well being responsible for the overall look and the story-boards).]

The type of situation ***Scampy*** is designed for is one in which events of some particular type are occurring over time, e.g., earthquakes, vehicles passing a checkpoint, or accidents. Raw data consists of the times at which the events occurred. There are two main ways of treating such data:

- to look at the distribution of the ***waiting times*** between the events
- to divide time up into intervals (e.g., hours, days, months), count the number of events occurring in each time interval, and look at the distribution of the ***numbers of events per time-interval***.

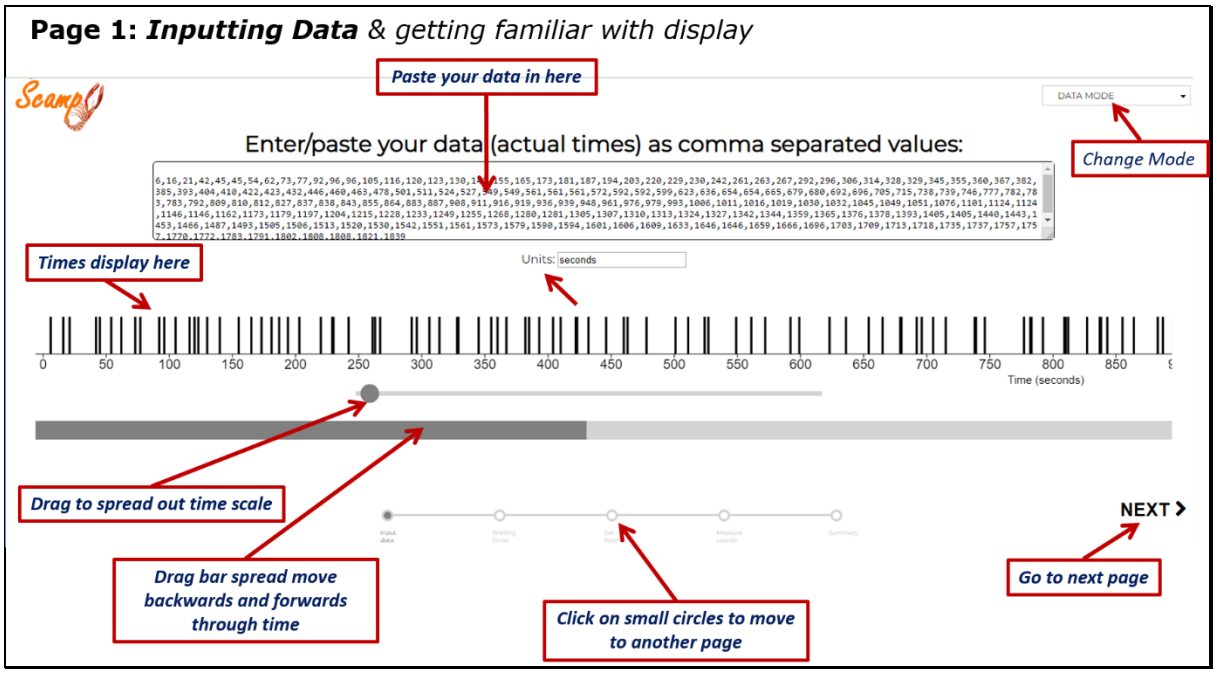
Under idealised conditions, the ***number of events per time interval*** has a **Poisson distribution** with mean λt. Here λ is the average occurrence rate and t is the length of a time interval. The distribution of the number of events is Poisson when the ***waiting times*** between events are a sample from an **Exponential distribution** with parameter λ (equivalently, with mean $1/\lambda$).

The Exponential distribution (https://en.wikipedia.org/wiki/Exponential_distribution) is not mentioned in either the curriculum or in Achievement Standard 3.14 (AS 91586) so this very nice Exponential/Poisson duality is not currently a central message at this level. But gaining a visual understanding of how randomness unfolds and how events-in-time data works is in scope. And the story-board pages that deal with the number of events per time interval and the Poisson distribution are core to 3.14.

Data-input modes

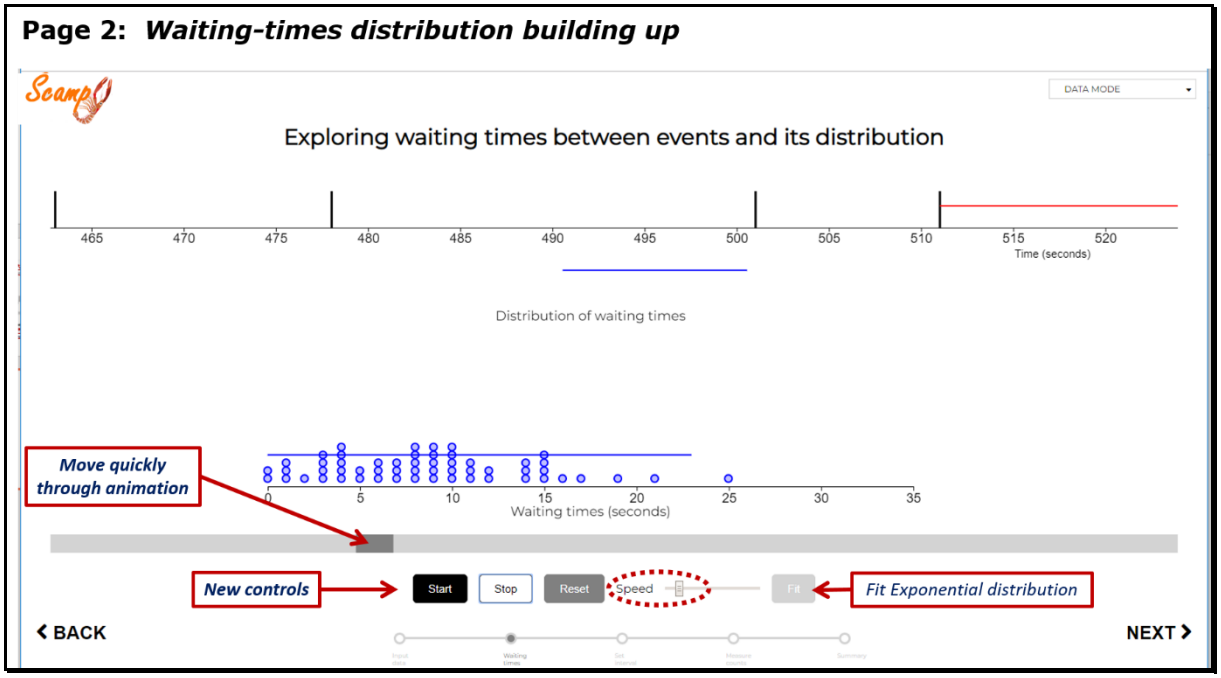
Scampy has a variety of ways of getting data to display: ("**Data Mode**") the user pastes in raw data; ("**Video Mode"**) the user get times by watching a video or listening to a sound file and pressing a button whenever the events occur; ("**Simulation Mode**") where the waiting times come from a user-selected *probability model*; or "**Game Mode**". Each mode allows for teachers to design ***different types of data-generating activities*** for students.

We'll start with the simplest mode, "**Data Mode**". A set of times at which some sort of event occurs is inputted by pasting a comma separated set of values into Page 1 of the story board (see below). A default set of times is already built in. To use your own, you paste them over the top of the ones that are there by default. (*Copy your data from a suitably formatted file, click into the enter-data box, Select All, and then paste on top of it*).



The app reacts by displaying all of the times along a horizontal time axis. This page also introduces the user to a number of controls. If the number of times is reasonably large, the times get too close together to see them individually. So there is a slider control to spread out time. This means that the entire time scale no longer fits on the page so there is another *slider-bar* control just beneath the first that enables you to move up or down the time axis. A time-line at centre-bottom lets you jump between story-board pages by clicking little buttons (this control is present on all pages).

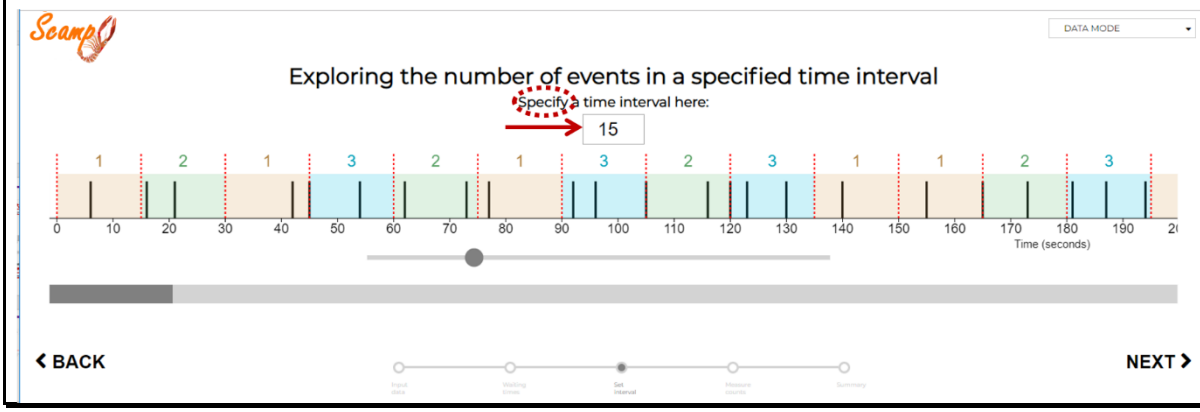
The point of the 1st page is gaining a feel for how the occurrence-time data looks using a very simple display before moving on to more sophisticated displays that are developed from this one.



On the 2nd page a dynamic display draws attention to the waiting times as they appear and then drops them down on to a horizontal axis for waiting-times. When they get near the waiting-times axis they turn into dots, thus building up a dot plot of the waiting times. There are controls for **Start**ing and **Stop**ping the

display and two ways to speed it up or slow it down. The **speed slider** essentially governs the play-back speed of a movie. The **slider bar** just under the dot plot gives much more hands-on control of the speed at which things happen (or even allows backing up the animation). The **Fit button** superimposes an exponential distribution with its mean set equal to the mean of the observed waiting times. It is a simple way of investigating whether the waiting times look as if they might have come from an exponential distribution.

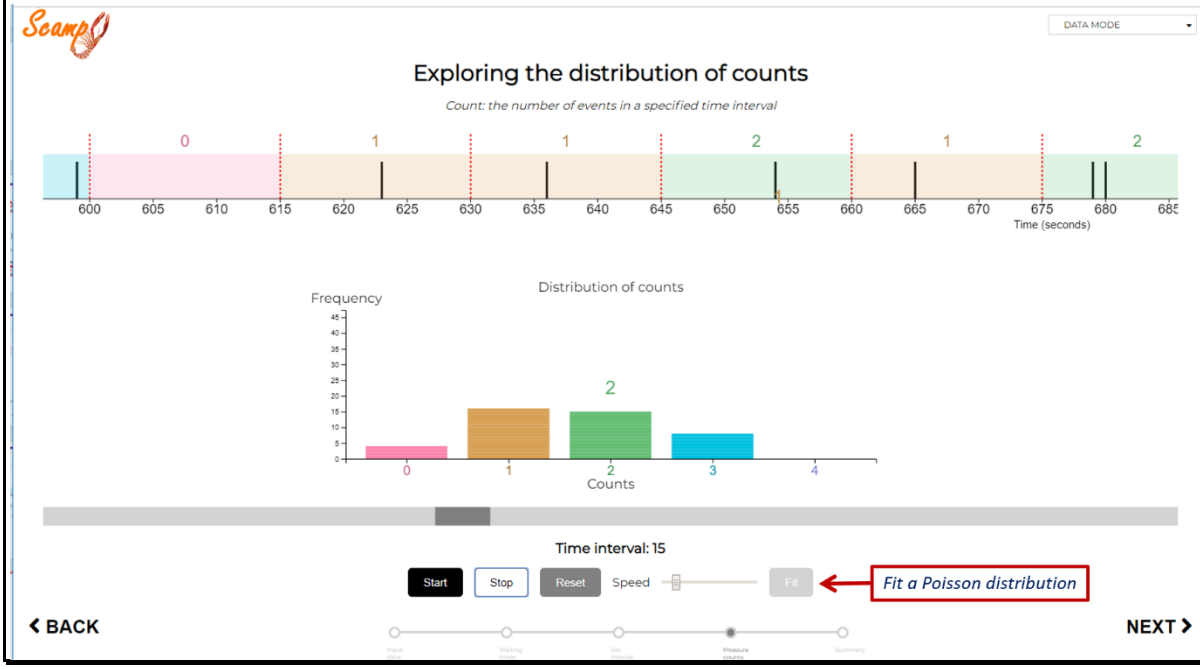
Page 3: *Number of events per time interval*



The point of the 3rd page is to start seeing things in terms of the numbers of events in a time interval and seeing the sequence of such counts obtained from the data.

Before this can begin the user has to type in the interval-width of the time intervals to be used (15 in the picture above). Red dotted lines are drawn with the selected spacings starting from zero. A ***count*** of the number of events appears for each time interval (the interval between two red dotted lines). The counts are colour-coded for reasons that will become apparent on the next page. We again have a slider to allow spreading out time so individual times can easily be seen, and a bar slider for moving up and down the time axis so that the whole sequence of counts can be seen.

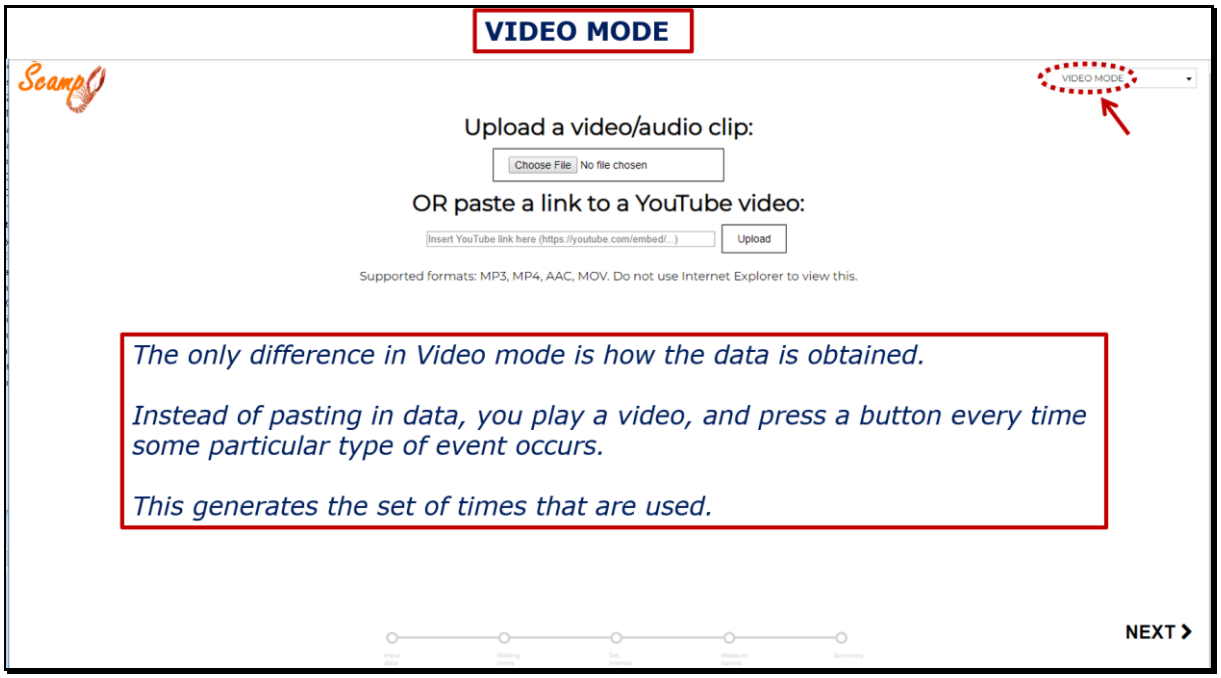
Page 4: *Counts distribution building up*



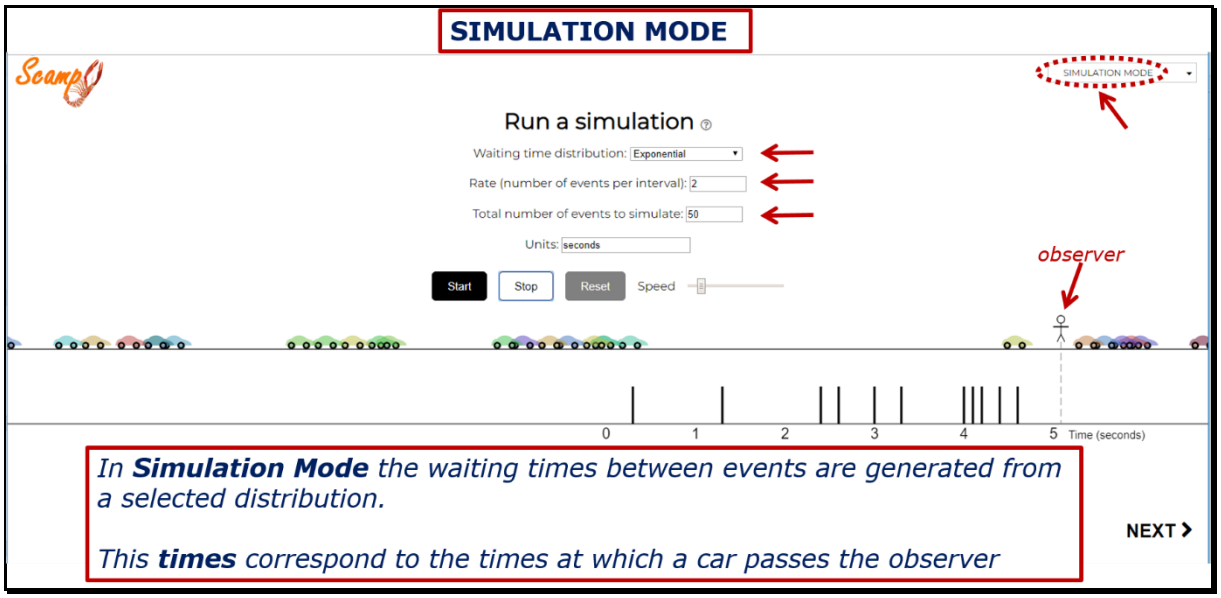
On the 4th page the scale starts drifting to the left and, as it is encountered, each count drops down onto an axis where a bar chart of the sequence of counts starts to build up. The counts are coloured with the same

colours as the bars where they end up to make it easier to track what is going where. Again we have a movie-speed slider and a bar-slider for hands-on control of the speed of which things happen (or even to back up the animation). Clicking the Fit button results in a Poisson distribution to be superimposed over the bar chart. The displayed Poisson is the one with its mean set equal to mean of the events-per-interval counts.

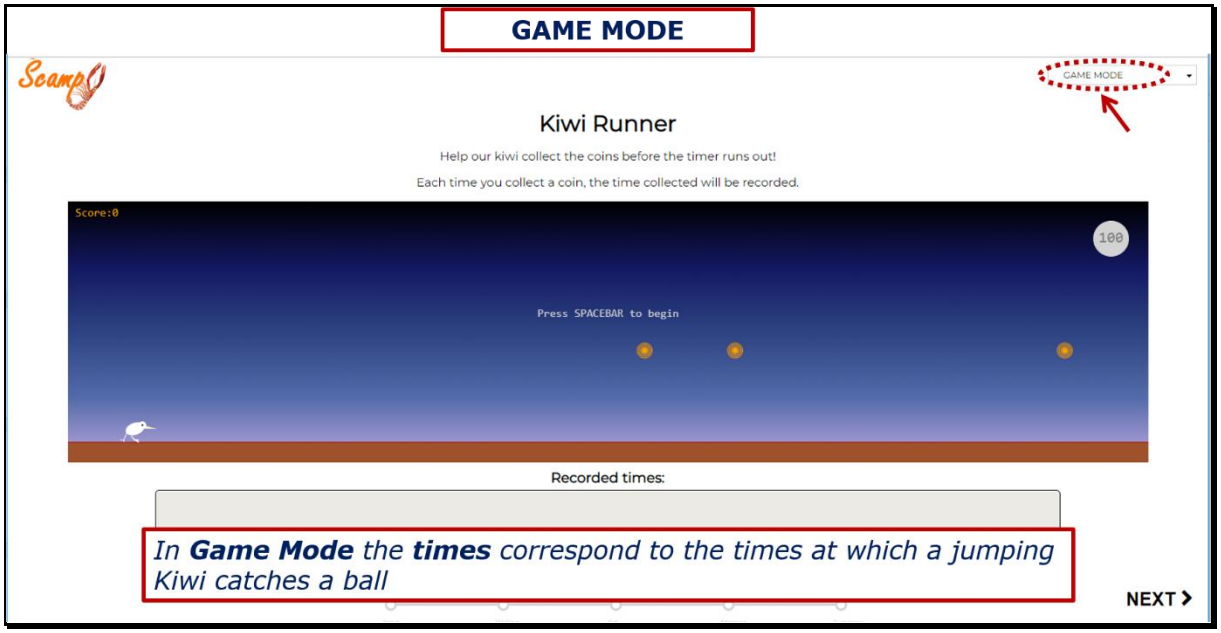
The 5th page shows the waiting-times distribution and also the event-counts distribution on the same page. The Fit button superimposes the exponential distribution over the waiting times and the Poisson over the event counts. This is motivated by the theoretical relationship between exponential waiting times and Poisson counts.



In **Video Mode**, once the video or sound file is loaded, it can be played in the page. A button appears that the viewers/listeners can use to press whenever the type of event they are looking for occurs in the video/sound track. The data that is used thereafter is the set of times the button was pressed. This makes for a classroom-friendly analogue to observing events in the wild.



In **Simulation Mode**, waiting times are generated from a user-selected distribution. The parameters offered depend on the distribution chosen. These waiting times determine the times at which the little cars in the image above pass the observer. A beep also sounds each time one passes. The resulting sounds give an alternative way of experiencing randomness through time. The simulation appears on each of the pages above.



Game Mode at resent is at present very simple but may be an interesting idea for development.