1.010 Uncertainty in Engineering
Fall 2008

Application Example 6

(Exponential and Poisson distributions)

ARE THE SEQUENCES OF BUS AND EARTHQUAKE ARRIVALS POISSON?

The Poisson Process

The Poisson process is the simplest random distribution of points on a line. What makes this model simple and convenient to use is the fact that, in a Poisson process, the number and locations of events in non-overlapping (separate) intervals are independent. This condition is often referred to as "lack of memory" of the process, since it implies that the pattern of past events has no relevance to the pattern of future events.

For some applications, the use of the Poisson point process can be justified on theoretical grounds, but in most cases one must verify the plausibility of the Poisson assumption by comparing implications of that assumption with actual data. Two verifications are frequently made:

1. If the point process is Poisson, the distribution of the interarrival time, T, is exponential with mean value $1/\lambda$ where λ is the rate parameter of the process;

2. If the point process is Poisson, the number of events in an interval of duration D has Poisson distribution with mean value λD.

Next we illustrate the use of these validation techniques for the process of bus arrivals at a station. The data sets used for this example are synthetic, but they are representative of patterns one may observe in actual samples.

Three synthetic data sets have been generated, each containing 100 arrival times in minutes. The data sets are representative of conditions at three different bus stations: Data Set 1 refers to bus arrivals at Point A in Figure 1, which is close to the dispatch station. Data Set 2 is representative of conditions at Points B, which is far from the dispatch station and Data Set 3 is collected at Point C, in a downtown area. The arrival times are given in the attached tables and are displayed graphically in Figures 2 and 3.
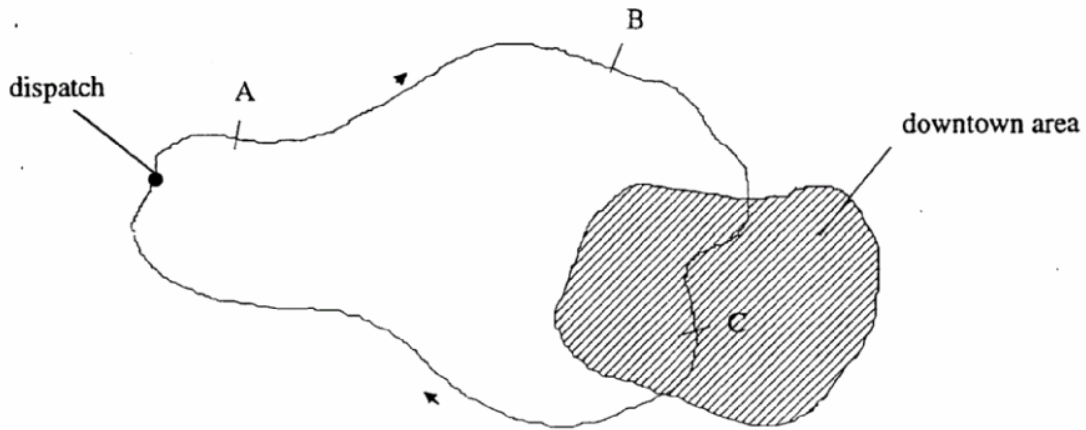

Figure 1: Locations of bus data collection station relative to dispatch point

7:03	7:07	7:11	7:15	7:20	7:24	7:25	7:28	7:33	7:43
7:46	7:47	7:49	7:51	7:53	7:58	8:04	8:10	8:16	8:18
8:20	8:25	8:28	8:33	8:35	8:37	8:44	8:50	8:58	9:01
9:09	9:14	9:15	9:18	9:23	9:27	9:29	9:34	9:38	9:40
9:48	9:51	9:55	10:00	10:06	10:10	10:15	10:17	10:24	10:27
10:31	10:38	10:43	10:45	10:51	10:54	10:57	10:59	11:01	11:11
11:14	11:16	11:21	11:28	11:32	11:37	11:45	11:48	11:51	12:01
12:04	12:13	12:16	12:21	12:26	12:29	12:38	12:45	12:51	12:55
13:00	13:06	13:15	13:21	13:24	13:26	13:28	13:31	13:35	13:42
13:47	13:53	13:55	14:02	14:06	14:08	14:11	14:16	14:21	14:28

Table 1: Arrival times at Station A

7:00	7:11	7:15	7:17	7:26	7:29	7:30	7:32	7:34	7:37
7:39	7:41	7:46	7:54	7:57	8:12	8:15	8:24	8:46	8:46
8:50	9:00	9:01	9:10	9:20	9:31	9:33	9:41	9:43	9:49
9:49	10:04	10:04	10:06	10:17	10:17	10:23	10:26	10:31	10:31
10:38	10:44	10:44	10:51	10:52	10:52	10:58	11:02	11:03	11:06
11:11	11:13	11:19	11:19	11:29	11:35	11:39	11:40	11:57	11:57
11:58	11:58	12:04	12:10	12:11	12:16	12:20	12:28	12:28	12:29
12:30	12:35	12:50	13:04	13:12	13:13	13:16	13:18	13:27	13:31
13:35	13:36	13:38	13:56	13:56	13:56	14:06	14:11	14:11	14:16
14:16	14:17	14:19	14:19	14:24	14:25	14:26	14:28	14:31	14:38

Table 2: Arrival times at Station B

| 7:01	7:01	7:01	7:29	7:29	7:42	7:50	7:58	7:58	8:00
8:00	8:10	8:13	8:41	8:41	8:47	9:15	9:16	9:24	9:31
9:31	9:31	9:31	9:36	9:36	9:59	10:01	10:01	10:01	10:03
10:03	10:18	10:18	10:18	10:18	10:21	10:24	10:24	10:25	10:25
10:25	10:25	10:48	10:51	10:51	10:52	11:51	11:52	11:52	11:52
11:52	11:52	12:05	12:15	12:24	12:24	12:25	12:32	12:39	12:40
12:40	12:46	12:47	12:58	13:00	13:00	13:05	13:07	13:07	13:07
13:07	13:07	13:07	13:33	13:36	13:36	13:36	13:36	13:36	13:36
13:36	13:36	13:36	13:39	13:39	13:39	13:39	13:39	13:39	13:39
13:39	13:39	13:43	13:43	14:00	14:14	14:17	14:19	14:19	14:20

Table 3: Arrival times at Station C
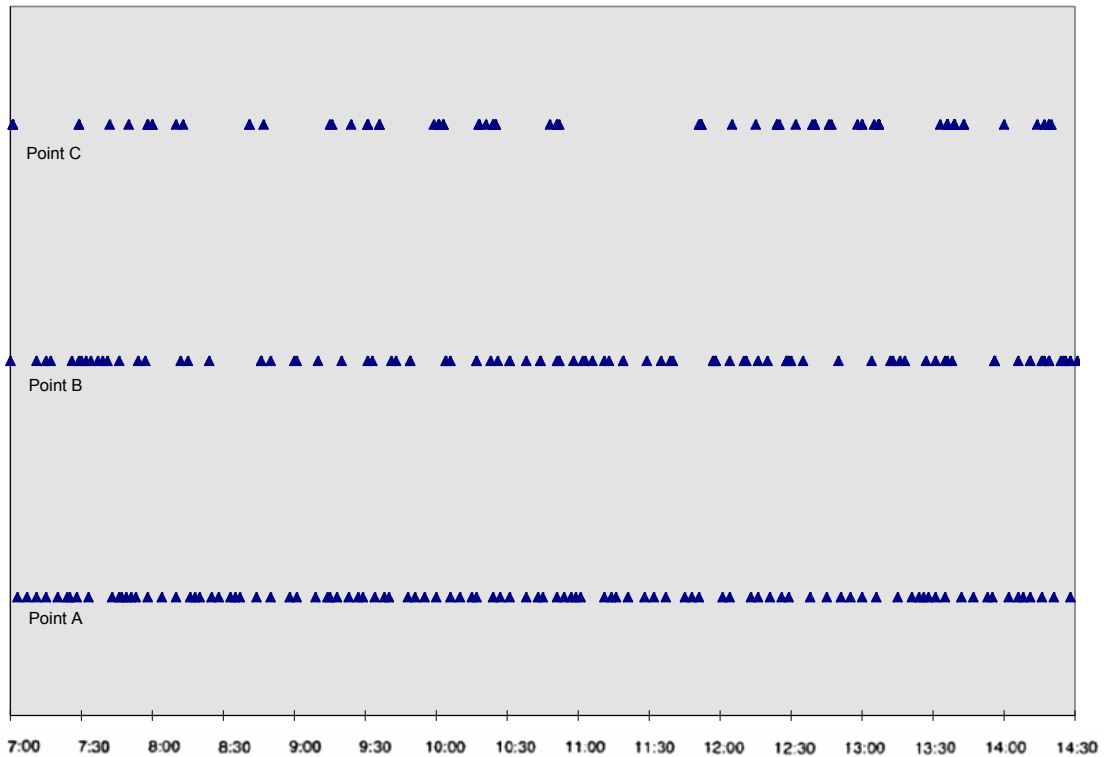
Figure 2: Pattern of arrival times at Stations A, B, and C
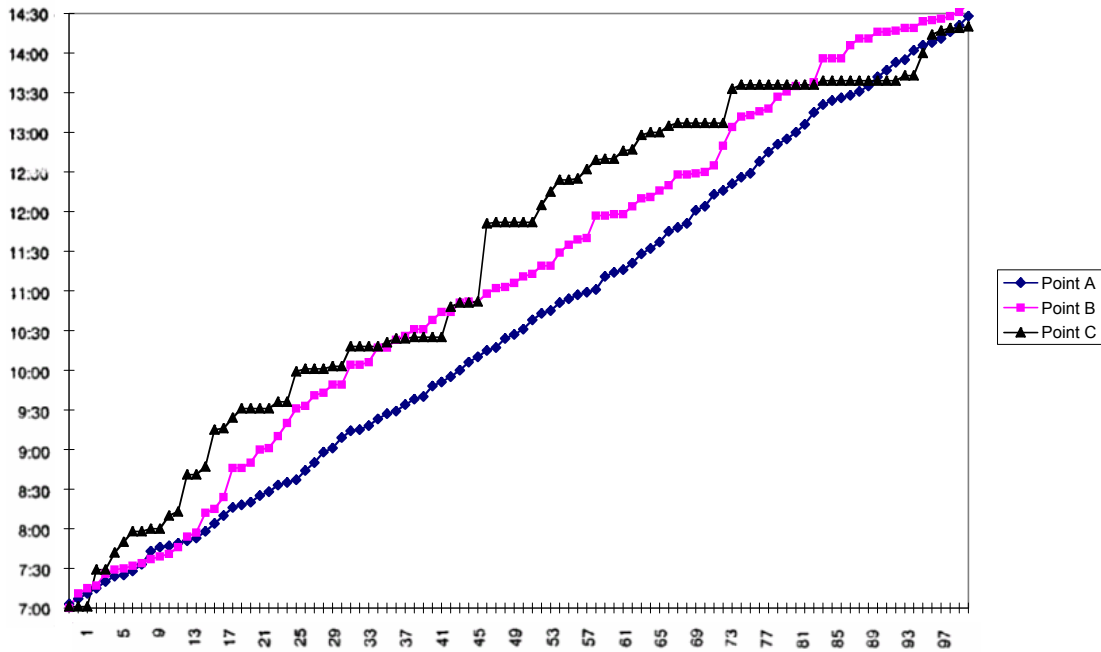
3

Figure 3: Arrival times of 100 buses

A cursory look at Figure 2 reveals differences in the three patterns. These differences become especially clear if one plots histograms of the interarrival time and compares such histograms to exponential probability density functions with the same mean values. These comparisons are made in Figures 4, 5 and 6.
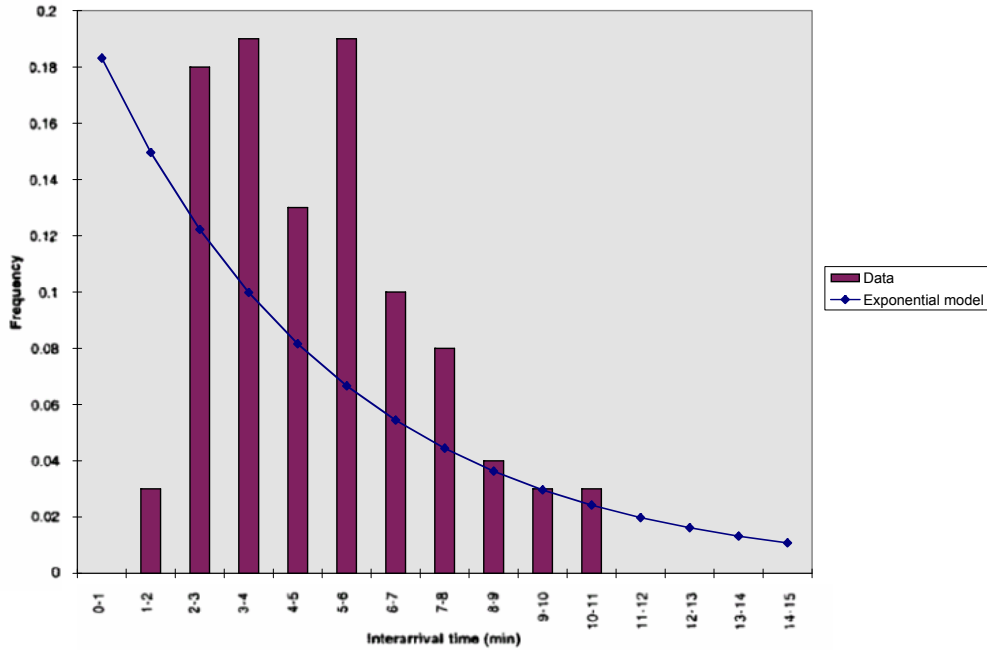
Figure 4: Histogram of the interarrival time at Station A; comparison with the exponential distribution
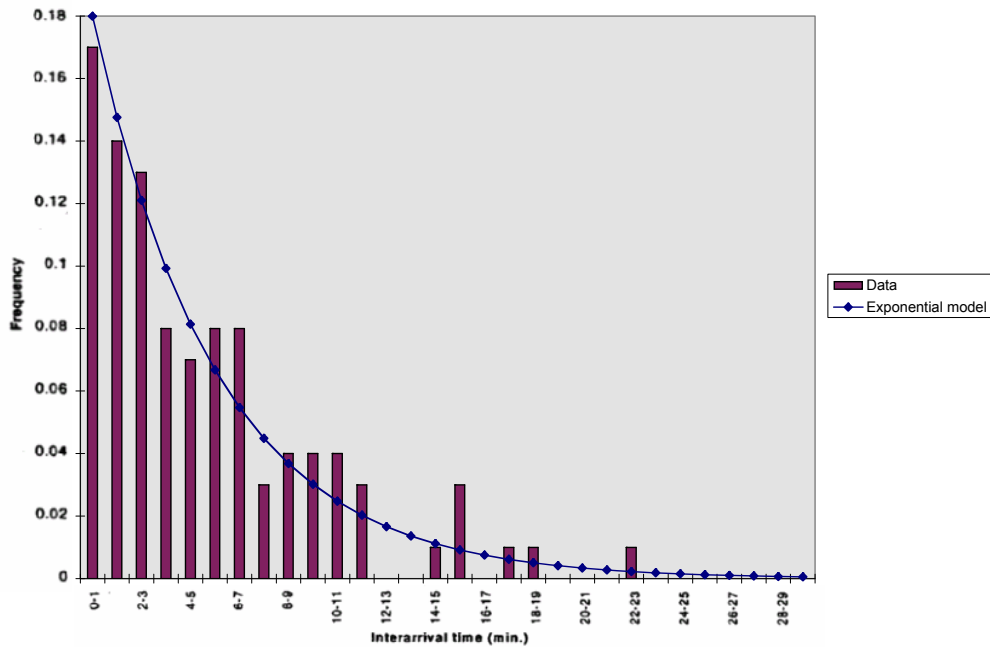

Figure 5: Histogram of the interarrival time at Station B; comparison with the exponential distribution
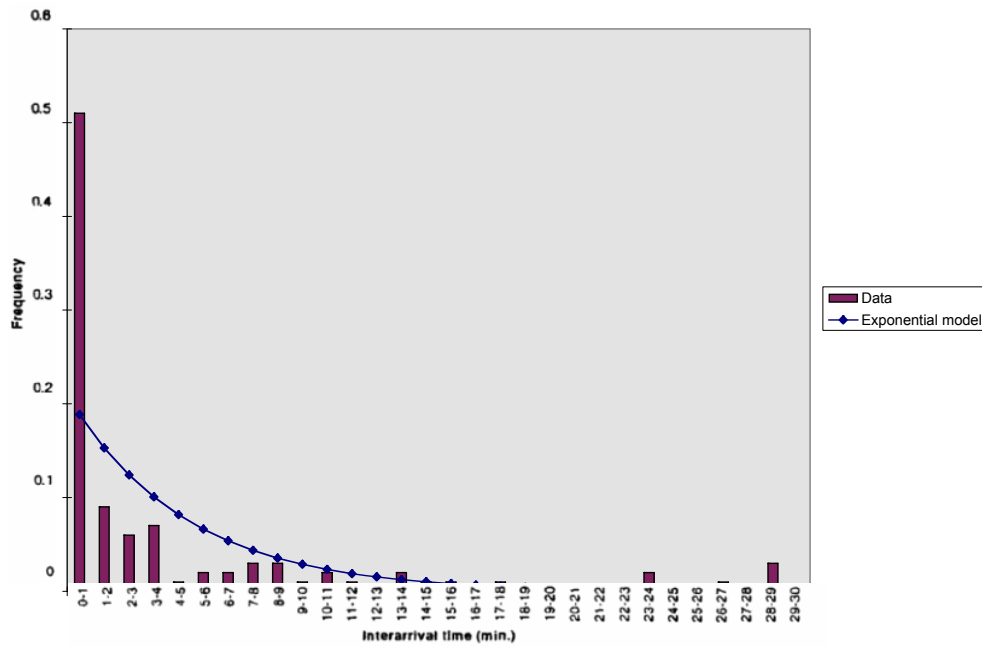
Figure 6: Histogram of the interarrival time at Station C; comparison with the exponential distribution

Notice that:

- All three samples have mean values close to 5 minutes, which is the theoretical mean value used in the simulation.

- At Station A, comparison of the histogram with the exponential distribution shows that, relative to a Poisson process, the data includes too many interarrival times close to the mean and too few very small or very large values. This is indicative of regularity in the process, i.e. the tendency of events to occur at nearly constant time lags. The reason why a record of this type may be representative of Station A is that, if buses are dispatched at equally spaced times, then their arrivals at stations close to the dispatch area tend to be regular.

- At Station B, the histogram has a nearly exponential decay, which is consistent with the Poisson assumption. The "randomization" of the times has occurred due to the

many random disturbances to the bus speed encountered from the dispatch point to Station B.

- At Station C, the histogram departs from the exponential shape in a way opposite to Station A. Specifically, relative to a Poisson process, Sample 3 contains too many very short and very long interarrival times and too few intermediate interarrival times. This pattern is symptomatic of clustering, which is the tendency of points to occur in widely separated groups. For buses, such condition may be induced by varying traffic congestion along the route or by the tendency under certain traffic conditions for vehicles to form platoons behind slowly moving vehicles.

Similar conclusions are arrived at by looking at the distribution of N_T, the number of bus arrivals in a period of fixed duration T. This is done in Figures 7, 8, and 9, for T = 15 min. Under the Poisson assumption, N_T has Poisson distribution, with mean value 3. Figure 7 displays an excessively high probability of 3 arrivals relative to the Poisson probability value. This is consistent with the previous observation that the arrival process at Station A is regular. Figure 8 shows good agreement between the empirical histogram and the Poisson distribution and Figure 9 displays deviations that are symptomatic of clustering (either very few or very many arrivals in T).
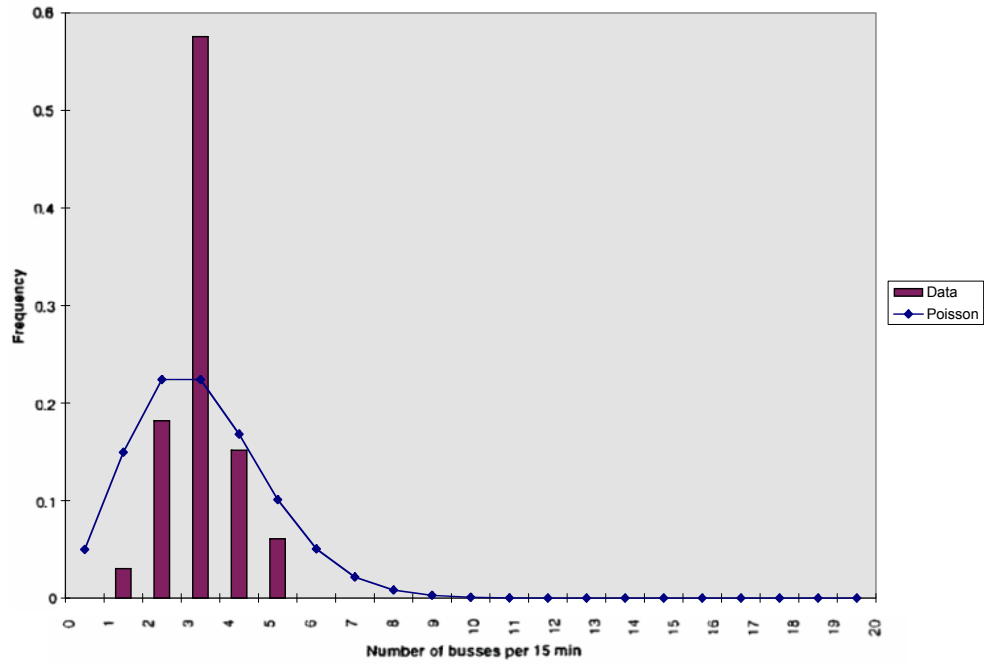
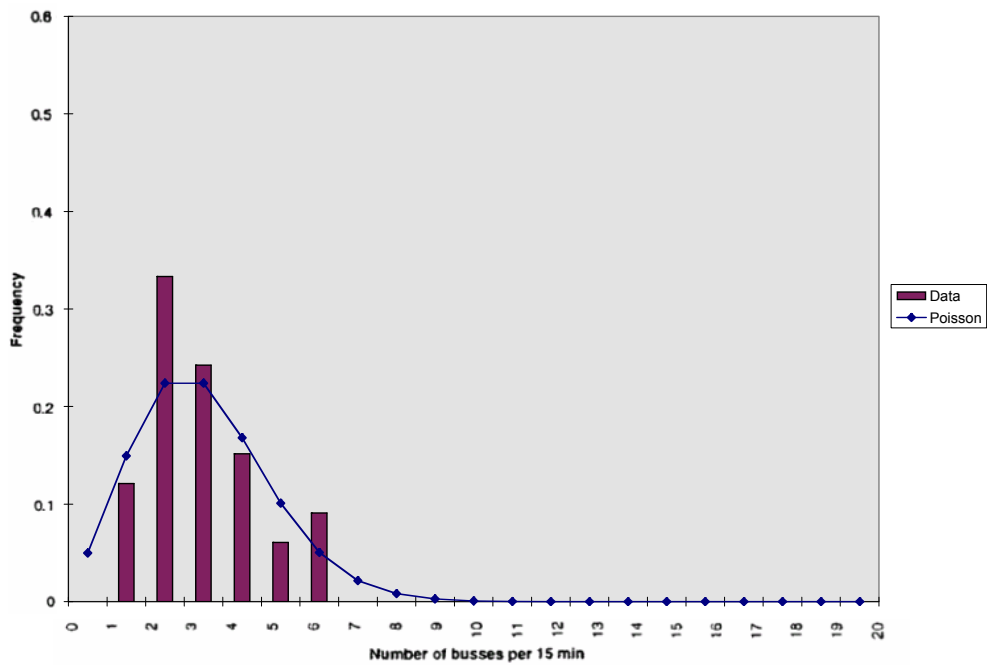**Figure 7: number of buses in periods of 15 minutes at Station A**

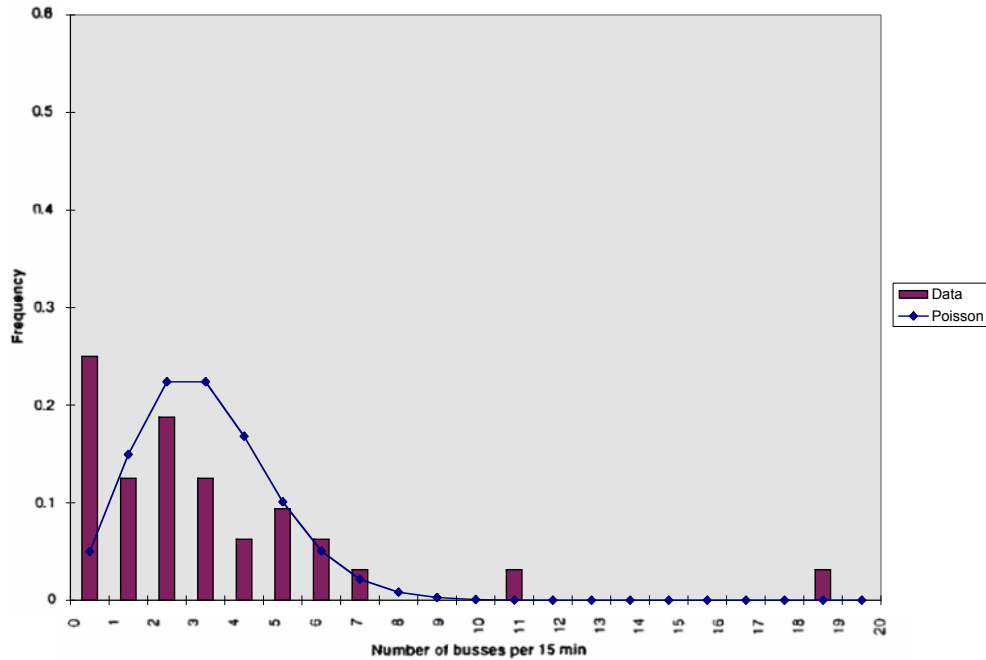**Figure 8: number of buses in periods of 15 minutes at Station B**

Figure 9: number of buses in periods of 15 minutes at Station C

Renewal Processes

The sequences of arrival times analyzed above were generated using a class of point processes called **renewal processes**. These processes are such that the interarrival times are independent and identically distributed. The Poisson process is a particular type of renewal process, with an exponential interarrival time distribution.

Specifically, the simulations of bus arrivals we made using a gamma distribution of the interarrival time, with shape parameter n = 5, 1, and 0.25 for Data set 1, 2, and 3, respectively. For n = 1, the gamma distribution is the same as the exponential distribution. Therefore, Data set 2 is indeed a sample from a Poisson process. Renewal processes are among the simplest processes through which one can represent regularity or clustering.

In nature, clustering is observed very frequently, because in many cases the occurrence of an event facilitates the occurrence of other events. An example is given by earthquake sequences, which display clustering due to foreshocks and aftershocks, swarms, etc.

9

Regularity is also observed, often as a result of competition or of a regulatory mechanism. For example, the times when airplanes take off or land at an airport tend to be regular, due to the minimum spacing imposed for safety by the traffic controllers. Similarly, the times at which customers leave a store tends to be regular if the service time does not vary much from customer to customer and the store has few counters.

In some cases, both clustering and regularity are observed in a point process, at different "scales". For example, it has been suggested and in some cases observed that earthquakes occur in clusters, as mentioned above, but the clusters themselves tend to be regularly spaced. Regularity of the clusters (more precisely, of the clusters that contain large magnitude earthquakes) is explained by the fact that large earthquakes "discharge" much of the elastic energy accumulated in the earth crust in the vicinity of the causative fault and it takes time for such energy to build up and reach again the critical level.

While any departure from the Poisson assumptions requires more complicated models to represent the phenomenon of interest, it is precisely the non-Poisson character of a point process that makes it possible to make predictions based on past observations. For example, if buses arrive rather regularly, say according to data set 1 and upon arriving at the station you are told that a bus just left, then you may "go for a cup of coffee" since it is very unlikely that another bus will be coming in the next few minutes. On the other hand, if the process is of the clustered type, then you should take the same information as an indication that another bus is likely to arrive soon. In the Poisson case, which is the dividing case between clustered and regular processes, you should be indifferent to the time since the last bus left.

Whether the sequence of major earthquakes from a given seismic source is Poisson, clustered, or regular is a hotly debated issue, with implications on the level of earthquake safety. For example, New England experienced the last strong earthquake in 1755. Based on the strain energy recharge theory and its regularity implications, some seismologists believe that we are now "due" for another similar earthquake, while other seismologists believe in the Poisson model and say that the fact that 250 years have passed without a major event is irrelevant to the future occurrence of large earthquakes in this area.

Problem 6.1

The table below gives the historical sequence of earthquake events in a region around the Messina Strait, which divides the island of Sicily from mainland Italy. The longest suspension bridge in the world is being planned to cross the Messina Strait. The region is highly seismic; hence a major concern is the threat posed by earthquakes to the bridge. In 1908, a disastrous earthquake struck the region, causing widespread damage and thousands of fatalities. The earthquake list provided in the table is limited to earthquakes of magnitude 4.5 or above, during the period 1700-1980. "Dependent" events, such as foreshocks and aftershocks, have been removed. Although this operation tends to make the resulting earthquake sequence "more Poissonian", certain dependences among the events usually remain.

  (a) Make an analysis of the interarrival times similar to that of bus arrivals in Fig. 4,

      (i) for all events irrespective of magnitude and

      (ii) only for events with magnitude greater than 5.

    Comment on the results.

  (b) Evaluate the probability that at least one event of magnitude larger than 5 occurs in the region during the 50 years from 1980 to 2030. Justify your estimate.

11

Year*	magnitude**	Year	magnitude
1702.13	4.5	1869.91	4.9
1707.17	4.5	1870.76	6.4
1712.54	4.5	1873.69	4.5
1716.14	4.5	1876.70	4.5
1717.30	5.6	1877.38	4.5
1720.69	4.5	1877.39	4.5
1720.70	4.9	1883.30	4.9
1729.49	4.5	1886.10	4.5
1736.62	4.9	1886.18	4.9
1739.36	5.6	1886.26	4.5
1743.18	4.5	1887.92	5.3
1743.93	4.9	1889.41	4.5
1767.53	5.6	1894.14	4.9
1770.00	4.5	1894.88	5.1
1783.09	6.9	1902.30	4.5
1783.10	4.5	1905.69	7.0
1786.18	5.6	1907.81	5.9
1789.10	4.5	1908.94	4.5
1791.78	5.6	1908.99	7.1
1792.36	4.5	1913.49	5.6
1805.52	4.9	1917.45	5.3
1806.27	4.7	1932.39	4.9
1821.58	4.9	1940.32	4.5
1823.17	6.0	1947.57	6.0
1824.94	4.9	1950.27	4.5
1828.19	4.9	1967.68	4.5
1831.08	4.9	1973.28	4.7
1832.18	6.8	1975.04	4.7
1835.78	5.6	1975.61	4.6
1836.31	6.0	1976.26	4.8
1836.34	4.5	1977.97	4.7
1852.06	4.5	1978.19	4.9
1852.07	4.5	1978.29	5.5
1854.11	5.6	1980.94	4.6

* time of occurrence is given here in fractional year

**magnitude has not been recorded directly (other than for a few recent events). The values reported here have been estimated from Modified Mercalli Intensity, which is a discrete scale. This is why the magnitudes tend to have discrete values.

Table 4: Historical catalog of "main events" (dependent events excluded) for a region surrounding the Messina Straights in Southern Italy. The catalog includes events of magnitude at least 4.5 for the period 1700-1980.