# Chapter 20

# Arrivals and waiting times

The arrival times of events in a Poisson process will be continuous random variables. In particular, the time between two successive events, say event $n-1$ and event $n$, we can denote by $T_n$. This is called the *interarrival time* or *waiting time*. In this notation, the waiting time for the first event/arrival is $T_1$

We can also define the total time until the $n$th event as $S_n$. Then in terms of the interarrival times we have:

$$S_n = T_1 + T_2 + \cdots + T_n \qquad . \tag{20.1}$$

By the same token we have:

$$T_n = S_n - S_{n-1} \quad n \geq 1 \qquad , \tag{20.2}$$

with the convention $S_0 = 0$.

## 20.1 Interarrival distributions

So if we have a series of 3 events which arrive at the following times (in minutes in decimal form): $2.11, 4.30, 4.80$. Then:

$$S_1 = 2.11 \quad , \quad S_2 = 4.30 \quad , \quad S_3 = 4.80 \qquad .$$

$$T_1 = 2.11 \quad , \quad T_2 = 2.19 \quad , \quad T_3 = 0.50 \qquad .$$

Consider, then, the first arrival, for which $S_1 = T_1$. Then $P(T_1 > t)$, which is the probability that the first event has not arrive, on or before a time, $t$, is the same as the probability that *no* events have occurred in the time interval $[0, t]$. That is:

$$P(T_1 > t) = P(N(t) = 0) \qquad . \tag{20.3}$$

That is, for a Poisson process with rate $\lambda$

$$P(T_1 > t) = e^{-\lambda t} \qquad . \tag{20.4}$$

It follows that:

$$P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - e^{-\lambda t} \qquad . \tag{20.5}$$

But this is nothing but the probability distribution function. Furthermore, it is the well-known exponential distribution. Therefore, for a Poisson process, the waiting time for the first arrival/event is exponentially distributed. This is consistent with the idea that both the Poisson process and exponential processes are *memoryless* distributions.

In a similar manner, one can deduce the distribution for any interarrival time $T_n$:

$$P(T_n > t) = P(N(t + S_{n-1}) - N(S_{n-1}) = 0) = P(N(t) = 0) = e^{-\lambda t} \qquad . \tag{20.6}$$

Then, again, we have:

$$F_{T_n}(t)) = 1 - e^{-\lambda t} \qquad , \quad n = 1, 2, 3, \ldots \qquad . \tag{20.7}$$

That is *all* waiting times have identical distributions (exponential), and as we have already seen:

$$\mathbb{E}(T_n) = \frac{1}{\lambda} \qquad , \qquad \mathrm{var}(T_n) = \frac{1}{\lambda^2} \qquad . \tag{20.8}$$

Moreover, these waiting times are also independent since they correspond to disjoint time intervals. Waiting times begin once the previous waiting time has ended!

In the previous chapter, the Poissonian nature of football was considered. probability distribution of goals per mach and the waiting times between goals were analysed. A total of 232 matches were taken as the data set and data for the 574 goals and the interarrival times. Note that matches were joined together (end-to-end) so that the total observation time was $232 \times 90$ minutes. In this case the average time between goals from the sample was 36.25 minutes and the standard deviation was 36.68 minutes. We recall that the (theoretical) exponential distribution is such that the mean and standard deviation are the same since:

$$\mathbb{E}(T) = \frac{1}{\lambda} \qquad , \qquad \mathrm{var}(T) = \frac{1}{\lambda^2} \qquad . \tag{20.9}$$

Moreover, we know the scoring rate is $\lambda = 2.478$ per match, that is $\lambda = 2.478/90$ per minute. According to the Poisson process this gives a mean interarrival time:

$$\mathbb{E}(T) = \frac{90}{2.478} \approx 36.3 \ \text{ mins} \qquad .$$

This value agrees well with the experimental data, which is very convincing evidence for a Poisson process. However, to confirm this is the case, we need to examine the distributions in detail. There is one difficulty using a continuous variable, sicne every goal will have its own exact time. To compare with theory we need to bin the times. In this case, Chu used 10-minute intervals. The results are shown in figure (**??**), along with the Poisson theoretical result which is the exponential distribution. Note that we allow the time to run past 90 minutes into the next game. However, in the data presented we gather all interarrival times greater than 130 mins into a single frequency. This explains why the distribution appears to go up towards the end.

The Poissonian behaviour of goal scoring is puzzling to a football supporter. There is a well-known saying in football that a team is most vulnerable to a goal being scored against them, immediately after they score themselves. In fact, this is true. The time of the next goal is given by an exponential distribution, and so the next goal is mostly like to occur in the 5 mins after the scoring of a goal (by either side).

But for an exponential process the following is also true. Given that a goal has not been scored in the 5 mins after the previous goal, it is *again* most likely that the next goal will be scored in the next 5 minutes. However, the same is true for a goal scored by either team. So your team is also most likely to score another goal in the next 5 minutes. So the vulnerability idea is a bit of a myth.

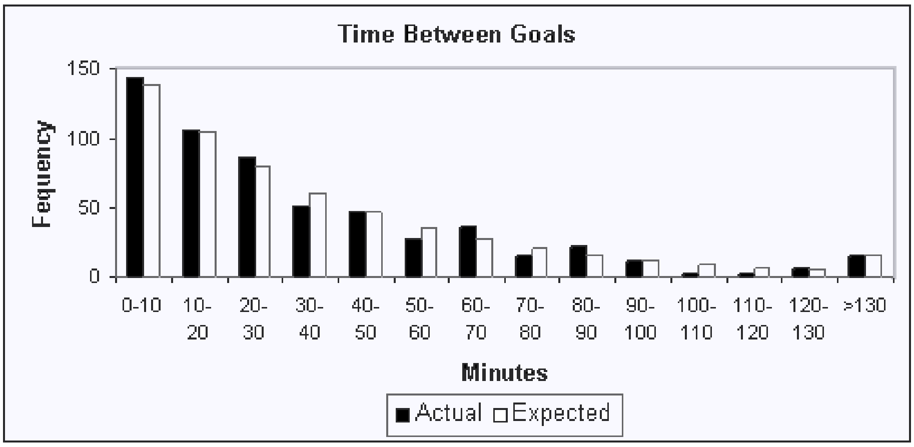


Figure 20.1: The sample distribution for waiting time for goals compared with an exponential distribution, with $\lambda = 2.478$ per game, that is $\lambda = 0.0275$ per minute. The data (black bars) agree well with the exponential model (white bars).

In reality, a few factors change the process. A football team is not Markovian, that is they possess memory in one simple sense, as the game progresses the get progressively more tired. So one cannot say the first half is similar to the second. Furthermore the team have awareness. Thus they are more likely to attack strongly in the last 10 minutes if they are losing the game. They are also more likely to play aggressively when playing at home. It's complicated. We will discuss how this can be modelled as an *inhomogeneous* Poisson process.

## 20.2 Arrival distributions

We have discussed the distribution for the first arrival, and the interarrival times, and found that these are exponential in form. Attention now turns to the distribution for the $n$th arrival.Recall that this is defined by

$$S_n = \sum_{i=1}^{n} T_i = T_1 + T_2 + \ldots + T_n \qquad . \tag{20.10}$$

Prior to obtaining the probability density for $S_n$, we can immediately deduce the expected value and variance, since the waiting times are independent and identically distributed so we have that:

$$\mathbb{E}(S_n) = n\mathbb{E}(T_1) = \frac{n}{\lambda} \quad \text{and} \quad \operatorname{var}(S_n) = \frac{n}{\lambda^2} \tag{20.11}$$

Then the probability density is by definition:

$$f_{S_n}(s)ds = P(s \le S_n \le s + ds) \qquad . \tag{20.12}$$

So considering a small increment in time, $h$ we have: the probability that the $n$nth arrival occurs in this interval:

$$P(t < S_n \le t + h)$$

And this is just the probabiliy that exactly $n-1$ events have occurred in the time interval $[0, t]$ and that a further single event occurs in the interval $(t, t+h]$.

Since these two time intervals are disjoint and this is a Poisson process, these probabilities are independent.

$$P(t < S_n \leq t+h) = P\{N(t) = n-1\} \times P\{N(t+h) - N(t) = 1\} \qquad .$$

This gives us:

$$f_{S_n}(t)h = \frac{\mathrm{e}^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}[\lambda h + o(h)]$$

Now taking $h \to 0$, leaves us with:

$$f_{S_n}(t) = \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} \qquad t \geqslant 0 \tag{20.13}$$

This is a special case of the *Gamma distribution*. This has the mean given by:

$$\mathbb{E}(S_n) = \int_0^{+\infty} t \times \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}dt = \frac{1}{\lambda(n-1)!}\int_0^{+\infty} e^{-x}x^n dx \qquad . \tag{20.14}$$

This integral is a Gamma function with value $n!$. This yields the final result,

$$\mathbb{E}(S_n) = \frac{n}{\lambda} \tag{20.15}$$

but this result was previously obtained (20.11) by a much simpler route.

## 20.3 Conditional density

Suppose we know that a process in Poissonian, and we know that a single event occurred in the interval: $0 \leq t \leq \tau$. For example, suppose we heard that our football team won a match 1-0. One interesting question we could pose is, was it more likely that the goal was scored in the first half or second half?

The answer is given by considering the *conditional* probability density for a Poisson process. In formal terms, suppose we know a single event has occurred during a time interval $[0, \tau]$, what is the probability that it occurred before a time $t$. This can be expressed mathematically as:

$$P(T_1 \leq t | T_1 \leq \tau) \quad , \quad 0 \leq t \leq \tau \qquad . \tag{20.16}$$

This in turn can be written as:

$$P(N(t) = 1 | N(\tau) = 1) = \frac{P(N(t) = 1, N(\tau) = 1)}{P(N(\tau) = 1)} = \frac{P(N(t) = 1, N(\tau) - N(t) = 0)}{P(N(\tau) = 1)} \qquad . \tag{20.17}$$

Now the numerator involves two disjoint time intervals: $[0, t]$ and $[t, \tau]$, that is we have effectively cut the interval $[0, \tau]$ in two. Disjoint times are independent for a Poisson process. Therefore whatever happens in these two intervals are independent events. Thus we can write:

$$P(N(t) = 1, N(\tau) - N(t) = 0) = P(N(t) = 1) \times P(N(\tau) - N(t) = 0) \quad . \tag{20.18}$$

Since the process is homogeneous (stationary), and since $N(0) = 0$, we can then write:

$$P(N(\tau) - N(t) = 0) = P(N(\tau - t) - N(0) = 0) = P(N(\tau - t) = 0) \quad . \tag{20.19}$$

Thus:

$$P(T_1 \leq t | T_1 \leq \tau) = \frac{P(N(t) = 1)P(N(\tau - t) = 0)}{P(N(\tau) = 1)} \qquad . \tag{20.20}$$

This leads to the result:

$$P(T_1 \le t | T_1 \le \tau) = \frac{e^{-\lambda t}\lambda t e^{-\lambda(\tau - t)}}{e^{-\lambda\tau}\lambda\tau} = \frac{t}{\tau} \quad , \tag{20.21}$$

for the conditional probability distribution. This gives the conditional probability density:

$$f_{T_1|T_1\le\tau}(t) = \frac{1}{\tau} \quad . \tag{20.22}$$

We recognise that this conditional density is uniform. Thus an interesting property of the Poisson process, and the answer to our question, is that the event could have occurred at any time with equal probability (given it occurred). Thus, the goal is equally likely to have been scored in the first half or second half, given that we know a single goal was scored.

## 20.4 A worked example: squashed frogs

Suppose that traffic arrives as a Poisson process with a rate $\lambda$ and it takes a frog a time $f$ to cross the road. If a car passes while the frog is on the road assume there is a probability $0 \le p \le 1$ that the car will squash the frog, perhaps due to the finite width of the tyres. What is the probability that the frog survives?

Suppose that $p = 1$ (certain squashing). The first car that comes along will squash the frog, if it arrives before a time $f$. Therefore the probability that the frog will cross safely, $P(S)$, is the same as the probability that the first car arrives after a time $f$. Therefore the frog survives and crosses safely when no events occur before a time $f$, that is:

$$P(S) = P(N(f) = 0) = e^{-\lambda f} \quad . \tag{20.23}$$

Now consider the extreme case $p = 0$, that means that no cars will squash the frog. Then clearly:

$$P(S) = 1 \quad . \tag{20.24}$$

So when $0 < p < 1$, we expect the survival probability to lie between these limits:

$$e^{-\lambda f} < P(S) < 1 \quad . \tag{20.25}$$

### 20.4.1 Solution by thinning

We can calculate the answer for $P(S)$ by considering it as a *thinning* Poisson process. That is we can assume that a fraction $p$ of the cars are *squashing cars* with certainty, while the other fraction $(1-p)$ are non-squashing cars. So if $N(t)$ denotes the total number of cars that pass by a time $t$, then:

$$N(t) = N_s(t) + N_{ns}(t) \quad ,$$

where $N_s(t)$ is the number of squashing cars that have passed by on or before a time $t$. Then $N_{ns}(t)$ the number of non-squashing cars that passed by the same time. Then, both processes are Poisson in their own right.

The intensity of *squashing cars*, is just $\lambda p$, and the intensity of non-squashing cars $\lambda(1-p)$.

We can ignore the non-squashing cars since the frog is immune to those. So effectively we have a reduced rate of traffic of squashing cars. We can simply use the result (20.23) replacing $\lambda$ by $\lambda p$, and we get:

$$\boxed{P(S) = P(N_s(f) = 0) = e^{-p\lambda f}} \quad . \tag{20.26}$$

This certainly lies within the limits (20.25).

### 20.4.2 Differential equation solution

We can derive the same result in a slightly different way. Let's consider $P(S)$ by conditioning on what happens over a short time $h \ll f$. Let $s(t)$ be the probability the frog is still alive after a time $t$ and consider the probability that it will be alive after a short further time $h$. Then, let the event $A$ be a car arriving in this (very short) time $h$, and event $B$ that the car squashes the frog. We suppose that $A$ and $B$ are independent So in this short time $h$ the frog will have either (a) been squashed or (b) covered a further fraction $h/f$ of the width of the road. Consider, first the partitioning rule in the form, where $S'$ is the event that the frog is still alive at time $t + h$:

$$P(S') = P(S'|A^c)P(A^c) + P(S'|A, B^c)P(A, B^c) + P(S'|A, B)P(A, B) \quad . \tag{20.27}$$

Note that the last term will be zero, since $P(S'|A, B)$, the probability it is alive, given that a car has arrived and squashed the frog, will be zero. The entire equation (20.27) can be written in this notation as:

$$s(t + h) = s(t)\,[1 - \lambda h + o(h)] + s(t)\,[\lambda h + o(h)]\,(1 - p) + 0 \quad . \tag{20.28}$$

We can convert this to a differential equation letting $h \to 0$:

$$\lim_{h\to 0} \frac{s(t + h) - s(t)}{h} = \frac{ds(t)}{dt} = -\lambda p s(t) \quad . \tag{20.29}$$

Separation of variables and integration gives:

$$\int \frac{ds}{s} = -\int \lambda p.dt \quad . \tag{20.30}$$

Given that $s(0) = 1$ we have the solution, at a time $f$:

$$\boxed{P(S) = s(f) = e^{-\lambda p f}} \quad . \tag{20.31}$$