1. Introduction

We will consider a process in which points occur randomly in time. The phrase **points in time** is generic and could represent, for example:

- The times when a piece of radioactive material emits particles
- The times when customers arrive at a service station
- The times when requests arrive at a server computer
- The times when accidents occur at a particular intersection

It turns out that under some basic assumptions that deal with independence and uniformity in time, a *single*, one-parameter probability model governs all such random processes. This is an amazing result and because of it the **Poisson process** (named after Simeon Poisson) is one of the most important in probability theory.

Random Variables

There are two collections of random variables that can be used to describe the process, corresponding to two different experiments; these collections are dual to one another in a certain sense.

First, let T_k denote the time of the k'th arrival for $k = 1, 2, ...$ The **gamma experiment** is to run the process until the k'th arrival occurs and note the time of this arrival. Next, let N_t denote the number of arrivals in the interval $(0, t]$ for $t \geq 0$. The **Poisson experiment** is to run the process until time t and note the number of arrivals. Note that

$$N_t \geq k \text{ if and only if } T_k \leq t$$

since each of these events means that there are at least k arrivals in the interval $(0, t]$.

The Basic Assumption

The assumption that we will make can be described intuitively (but imprecisely) as follows: If we fix a time t, whether constant or one of the arrival times, then the process *after* time t is independent of the process *before* time t and behaves probabilistically just like the original process. Thus, the random process has a **regeneration property**. Making this assumption precise will enable us to derive the distribution of each of the following in turn:

- The interarrival times
- The arrival times
- The number of arrivals in an interval

1. Think about the basic assumption for each of the specific applications given above.

Virtual Laboratories > The Poisson Process > 1 [2] 3 4 5 6 7 8

2. The Exponential Distribution

The basic assumption for the Poisson process is that the behavior of the process after an arrival should be independent of the behavior before the arrival and probabilistically like the original process (regeneration).

The Interarrival Times

In particular, the general regeneration assumption means that the times between arrivals, known as **interarrival times**, must be independent, identically distributed random variables. Formally, the interarrival times are defined as follows:

$$X_1 = T_1,\ X_k = T_k - T_{k-1} \text{ for } k = 2, 3, \ldots$$

where T_k is the time of the k'th arrival. We will assume that

$$P(X_i > t) > 0 \text{ for each } t > 0.$$

Now, we also want regeneration to occur at a fixed time t. In particular, if the first arrival has not occurred by time t, then the time remaining until the arrival occurs has the same distribution as the first arrival time itself. This is known as the **memoryless property** and can be stated in terms of a generic interarrival time X as follows

$$P(X > t + s) \mid X > s) = P(X > t) \text{ for all } s, t \geq 0.$$

Distribution

Let G denote the right-tail distribution function of X:

$$G(t) = P(X > t),\ t \geq 0.$$

1. Show that the memoryless property is equivalent to the **law of exponents**:

$$G(t + s) = G(t)G(s) \text{ for all } s, t \geq 0.$$

2. Show that the only solutions of the functional equation in Exercise 1, which are continuous from the right, are exponential functions. Let $c = G(1)$. Successively show that

a. $G(n) = c^n$ if n is a positive integer.
b. $G(1/n) = c^{1/n}$ if n is a positive integer.
c. $G(m/n) = c^{m/n}$ if m and n are positive integers.
d. $G(t) = c^t$ for any $t > 0$.

In the context of Exercise 2 let $r = -\ln(c)$. Then $r > 0$ (since $0 < c < 1$) so

$$G(t) = P(X > t) = e^{-rt},\ t \geq 0.$$

Hence X has a continuous distribution with cumulative distribution function given by

$$F(t) = P(X \le t) = 1 - G(t) = 1 - e^{-rt},\ t \ge 0.$$

3. Show that the density function of X is

$$f(t) = re^{-rt},\ t \ge 0.$$

A random variable with this density is said to have the **exponential distribution** distribution with **rate parameter** r. The reciprocal $1 / r$ is known as the scale parameter.

4. Show directly that the exponential density really is a probability density function.

5. In the exponential experiment, vary r with the scroll bar and watch how the shape of the probability density function changes. Now set $r = 2$, run the experiment 1000 times with an update frequency of 10, and watch the apparent convergence of the empirical density function to the probability density function.

6. In the exponential experiment, set $r = 1$. Run the experiment 1000 times, updating after each run. Compute the appropriate relative frequencies to empirically investigate the memoryless property

$$P(X > 3 \mid X > 1) = P(X > 2).$$

7. Show that the quantile function of X is

$$F^{-1}(p) = -\ln(1 - p) / r \text{ for } 0 < p < 1.$$

In particular, the median of X occurs at $\ln(2) / r$, the first quartile at $[\ln(4) - \ln(3)] / r$, and the third quartile at $\ln(4) / r$.

8. Suppose that the length of a telephone call (in minutes) is exponentially distributed with rate parameter $r = 0.2$.

a. Find the probability that the call lasts between 2 and 7 minutes.
b. Find the median, the first and third quartiles, and the interquatile range of the call length.

9. Suppose that the lifetime of a certain electronic component (in hours) is exponentially distributed with rate parameter $r = 0.001$.

a. Find the probability that the component lasts at least 2000 hours.
b. Find the median, the first and third quartiles, and the interquatile range of the lifetime.

Moments

The following exercises give the mean, variance, and moment generating function of the exponential distribution.

10. Show that $E(X) = 1 / r$.

11. Show that $\text{var}(X) = 1 / r^2$.

12. Show that $E[\exp(uX)] = r / (r - u)$ for $u < r$.

The parameter r is known as the **rate** of the Poisson process. On average, there are $1 / r$ time units between arrivals, so the arrivals come at an average rate of r per unit time. Note also that the median is always smaller than the mean for the exponential distribution:

$$\ln(2) / r < 1 / r.$$

13. In the exponential experiment, vary r with the scroll bar and watch how the mean/standard deviation bar changes. Now set $r = 0.5$, run the experiment 1000 times with an update frequency of 10, and watch the apparent convergence of the empirical mean and standard deviation to the distribution mean and standard deviation, respectively.

14. Suppose that the time between requests to a web server (in seconds) is exponentially distributed with rate parameter 2.

a. Give the mean and standard deviation of the time between requests.
b. Find the probability that the time between requests is less that 0.5 seconds.
c. Find the median, the first and third quartiles, and the interquatile range of the time between requests.

15. Suppose that the lifetime X of a fuse (in 100 hour units) is exponentially distributed with $P(X > 10) = 0.8$.

a. Find the rate parameter.
b. Find the mean and standard deviation.
c. Find the median, the first and third quartiles, and the interquatile range of the lifetime.

16. The position X of the first defect on a digital tape (in cm) has the exponential distribution with mean 100.

a. Find the rate parameter.
b. Find the probability that $X < 200$ given $X > 150$.
c. Find the standard deviation.
d. Find the median, the first and third quartiles, and the interquatile range of the position.
