Stacking Exponential Distributions

Anthony Charles

Last revised: August 8, 2023

1 Exponential distributions

In probability theory, exponential distributions are used to model processes where events occur continuously and independently, with a fixed average rate of events occurring that is independent of time. Specifically, if we denote $t = 0$ as the current time, then the probability distribution function for an event occurring at a future time t is given by

$$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau} \ , \tag{1.1}$$

where τ is a free parameter that is sometimes referred to as the mean time to maneuver (MTTM). That is, τ is the average amount of time it takes for an event to occur. For a random variable T sampled from this exponential distribution, the expected value of T is simply τ:

$$\mathbb{E}[T] = \int_0^\infty t\, f(t;\tau) dt = \int_0^\infty \frac{t}{\tau} e^{-t/\tau} = \Big[-\tau e^{-t/\tau} \Big]_0^\infty = \tau \ . \tag{1.2}$$

So, in order to pick the next maneuver times for a bunch of particles, we first have to specify this MTTM parameter τ, and then simply take a sample of $T \sim f$ for each particle and set the particle's next maneuver time equal to the sampled value of T.

Crucially, the exponential distribution is "memoryless" in the sense that the waiting time for an event to occur is independent of how much time has previously elapsed. That is, if $T \sim f$, then

$$\mathbb{P}(T > t) = \mathbb{P}(T > s + t \,|\, T > s) \ , \tag{1.3}$$

for all values of the times s and t. (See section 5 for the proof of this property.) This means that the average rate at which events occur is independent of time, i.e. constant. We will use this fact heavily in the next section when discussing stacked exponential distributions.

2 Stacking exponential distributions

Let's consider a more general case where we have N different types of maneuvers possible for a particle to do, indexed by $i = 1, \ldots, N$, each with their own associated MTTM τ_i. We assume that these are *independent*, such that performing a maneuver of type i will not affect the probability of doing a maneuver of type $j \neq i$.

The inefficient way to handle this is to now have N different "next time to maneuver" parameters for each particle, one per type of maneuver, and sample from the appropriate exponential distributions to generate each of these. This increases the state space of the particle, so we'd like to avoid this. Instead, we want to "stack" these into a single probability distribution, sample from that, and upon hitting the next maneuver time for this "stacked" distribution, we pick one of the maneuver types i to do, weighted by the respective values of τ_i.

To accomplish this, we first note that maneuver i having a MTTM τ_i is equivalent to have a maneuver frequency of $\omega_i = 1/\tau_i$. That is, instead of saying that the average time between maneuvers of type i is τ_i, we instead say that the expected number of maneuvers of type i happening in a unit of time is $1/\tau_i$. Then, the *total* expected number of maneuvers of *any* type happening per unit time is the sum of these individual number of events per unit time:

$$\omega_0 = \sum_{i=1}^{N} \omega_i = \sum_{i=1}^{N} \frac{1}{\tau_i} \ . \tag{2.1}$$

This leads us to define our “stacked" probability distribution for maneuvers of *any* type to be given by an exponential distribution with MTTM τ_0 as follows:

$$\tau_0 = \frac{1}{\omega_0} = \left(\sum_{i=1}^{N} \frac{1}{\tau_i} \right)^{-1} \ . \tag{2.2}$$

That is, if we generate maneuvers from the exponential distribution $f(t;\tau_0)$, then the total number of maneuvers that are done over any period of time is (on average) equal to the total number of maneuvers done over all maneuver types i when generating each individually from $f(t;\tau_i)$.

Once we have used $f(t;\tau_0)$ to decide that a maneuver has happened, we now have to decide what kind of maneuver has happened. The probability that a maneuver is of type i is given by

$$(\text{Probability of maneuver type } i) = \frac{\omega_i}{\omega_0} = \frac{1}{\tau_i} \left(\sum_{j=1}^{N} \frac{1}{\tau_j} \right)^{-1} \ , \tag{2.3}$$

since the frequency ω_i is the number of maneuvers of type i that occur per unit time, and ω_0 is the total number of maneuvers of any type that occur per unit time. That is, the distribution of possible maneuvers is a simple point-mass distribution, where the weight of maneuver i is ω_i/ω_0 and the weights clearly sum to one. This point-mass distribution can be sampled using the usual method of generating a single random number $U \sim \text{Unif}[0,1]$ and using the usual cumulative sampling technique.

3 Derivation using rejection probability

A way to more formally derive the expression (2.2) for the mean time of the stacked distribution is to consider products of rejection probability distribution for each maneuver. First, we note that the CDF of the exponential distribution (1.1) is given by

$$F(t;\tau) = \int_0^t f(t';\tau)dt' = 1 - e^{-t/\tau} \ . \tag{3.1}$$

We can think of $F(t;\tau_i)$ as the probability that a maneuver of type i has occurred at some time at or before t. Equivalently, the probability that *no* maneuver of type i has occurred at or before t (i.e. the rejection probability distribution) is

$$R(t;\tau_i) = 1 - F(t;\tau_i) = e^{-t/\tau_i} \ . \tag{3.2}$$

The total rejection probability, i.e. the probability that no maneuver of any kind has occurred at or before t, is given by the product over all individual rejections, since the maneuvers are assumed to be independent of one another. So, we have

$$R_{\text{total}}(t) = \prod_{i=1}^{N} R(t;\tau_i) = \prod_{i=1}^{N} e^{-t/\tau_i} = \exp\left[-t \left(\sum_{i=1}^{N} \frac{1}{\tau_i} \right) \right] \ , \tag{3.3}$$

using the fact that $e^a e^b = e^{a+b}$. So, the total rejection probability distribution is exactly of the form (3.2), i.e. a function of the form $e^{-t/c}$ for some constant c. In particular, we find that

$$R_{\text{total}}(t) = R(t;\tau_0) = e^{-t/\tau_0} \ , \quad \text{with } \frac{1}{\tau_0} = \sum_{i=1}^{N} \frac{1}{\tau_i} \ . \tag{3.4}$$

This matches the MTTM we derived in equation (2.2) exactly. Moreover, we can now go from $R(t;\tau_0)$ to the CDF $F(t;\tau_0) = 1 - R(t;\tau_0)$, which tells us that the probability of maneuvering is precisely what we found in the previous section.

4 Two-maneuver example

Consider a system where an object is moving at a fixed velocity, but it will randomly change either its speed or its course over time, with these maneuvers described by exponential distributions. We denote the mean time for a speed maneuver as τ_S and the mean time for a cours maneuver as τ_C. Concretely, we'll choose them as

$$\begin{aligned} \tau_S &= 20 \text{ minutes} \ , \\ \tau_C &= 5 \text{ minutes} \ . \end{aligned} \tag{4.1}$$

This means that the frequency of speed maneuvers ω_S and frequency of course maneuvers ω_C are given by

$$\begin{aligned} \omega_S &= 0.05 \text{ speed maneuvers per minute} \ , \\ \omega_C &= 0.20 \text{ course maneuvers per minute} \ . \end{aligned} \tag{4.2}$$

In total, then, we expect 0.25 maneuvers per minute of any type. That is,

$$\omega_0 = \omega_S + \omega_C = 0.25 \text{ maneuvers per minute} \ . \tag{4.3}$$

So, the mean time to maneuver for any type of maneuver is given by

$$\tau_0 = \frac{1}{\omega_0} = 4 \text{ minutes} \ . \tag{4.4}$$

That is, we expect to see one maneuver (of either speed or course) every four minutes. When modeling the moving object with e.g. a particle filter, the "next time to maneuver" is determined by sampling from the exponential distribution $f(t;\tau_0)$. Once we have hit the time of a maneuver, we need to determine whether it's a speed maneuver or a course maneuver. The frequency of course maneuvers is four times that of speed maneuvers ($\omega_C = 4\omega_S$), and so the probability of deciding that we've done a course maneuver should be four times that of a speed maneuver. This is supported by the associated weight calculations:

$$\begin{aligned} (\text{Probability of speed maneuver}) &= \frac{\omega_S}{\omega_0} = \frac{0.05}{0.25} = 0.2 \ , \\ (\text{Probability of course maneuver}) &= \frac{\omega_C}{\omega_0} = \frac{0.20}{0.25} = 0.8 \ . \end{aligned} \tag{4.5}$$

We can sample $U \sim \text{Unif}[0,1]$ and determine the maneuver type by comparing it to these weights:

$$\begin{aligned} 0 \leqslant U < 0.2 &\implies \text{speed maneuver} \ , \\ 0.2 \leqslant U < 1 &\implies \text{course maneuver} \ . \end{aligned} \tag{4.6}$$

We therefore have a simple way to generate both speed and course maneuvers from a single stacked exponential distribution whose frequency is the sum of constituent frequencies.

5 Proving the memoryless property

Let T be a random variable distributed according to the exponential distribution $f(t;\tau) = \frac{1}{\tau}e^{-t/\tau}$. The corresponding CDF $F(t;\tau)$ is given in (3.1). The quantity $\mathbb{P}(T > t)$ is simply the rejection probability distribution (3.2) that we have already computed, since $T > t$ effectively means that the event has not occurred within time t. That is,

$$\mathbb{P}(T > t) = 1 - \mathbb{P}(T \leqslant t) = 1 - F(t;\tau) = e^{-t/\tau} \ . \tag{5.1}$$

With that out of the way, our goal now is to compute the quantity $\mathbb{P}(T > s + t \,|\, T > s)$. First, we recall that Bayes' theorem says that for any two events A and B, the probability of both events occuring can be broken up into conditional probabilities as $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Using this, we find that

$$\mathbb{P}(T > s + t \,|\, T > s) = \frac{\mathbb{P}(T > s + t \ \cap \ T > s)}{\mathbb{P}(T > s)} \ . \tag{5.2}$$

Of course, since we are setting the starting time to be zero and considering times $s, t \geqslant 0$, this means that $T > s + t$ automatically implies that $T > s$. So, the second condition in the numerator is trivial, and this simplifies to

$$\mathbb{P}(T > s + t \,|\, T > s) = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} \ . \tag{5.3}$$

Plugging in the rejection distribution (5.1), we find that:

$$\begin{aligned} \mathbb{P}(T > s + t \,|\, T > s) &= \frac{e^{-(s+t)/\tau}}{e^{-s/\tau}} \\ &= e^{-t/\tau} \\ &= \mathbb{P}(T > t) \ . \end{aligned} \tag{5.4}$$

This is the memoryless condition (1.3).