CHAPTER 6

Exponential Decay Revisited: A Probabilistic Theory

In both problems of radioactive decay and the HIV dynamics, we have exponential decay, of the population of radioactive nuclei and of the concentration of viral particles, that takes the form

$$X(t) = X(t_0)e^{-c(t-t_0)}, \tag{6.1}$$

in which the constant c is known as a decay rate. This behavior is a consequence of simple "death" events. At the most basic level, both problems are actually about counting numbers of individuals: The still radioactive nuclei and the infectious viral particles at time t. As we have discussed earlier, equations like (6.1) with real-valued, deterministic variable $X(t)$, as a model for discrete individuals with stochastic behavior, does not elicit much confidence in us toward the mathematics.

But there is a more fundamental reason for the origin of the widely observed exponential decay. It has to do with the concept of *independent rare event*: Something with a very small probability to occur.

To show this, let us consider the waiting time to win a jackpot, t. Let $P(t)$ being the probability you still *have not* won up to time t. Then

$$P(t + \Delta t) = P(t)P(\Delta t)$$

Here we have assumed that jackpots comes up all the time, with probabilities of winning being independent. [This is the same argument as that there is no difference for one's luck in picking a card the first or the last.]

We now further consider the probability of winning as the function of Δt when Δt is very small. It has to be true that when $\Delta t = 0$, the probability is zero. Therefore Taylor expansion tells us

$$1 - P(\Delta t) = \lambda \Delta t + o(\Delta t),$$

in which the Bachmann-Landau notation $o(x)$ means

$$\lim_{x \to 0} \frac{o(x)}{x} = 0.$$

See `http://en.wikipedia.org/wiki/Big_O_notation`. The λ is the rate of increasing probability, as time goes on, of winning the jackpot. With these assumptions, we therefore have

$$P(t+\Delta t) = P(t)\left[1-\lambda\Delta t - o(\Delta t)\right]. \tag{6.2}$$

That is

$$\begin{aligned} P(t+\Delta t) - P(t) &= -\left[\lambda\Delta t + o(\Delta t)\right]P(t), \\ \frac{P(t+\Delta t) - P(t)}{\Delta t} &= -\left[\lambda + \frac{o(\Delta t)}{\Delta t}\right]P(t), \\ \frac{dP(t)}{dt} &= -\lambda P(t). \end{aligned}$$

Therefore, the *probability of still not winning a jackpot* is exponentially decreasing. We also see that almost any event is "rare" in the face of "continuous time": Some events might not be rare on the order of days, but it will be in terms of femtosecond (10^{-15})!

This is a probabilistic theory about any event. Now applying it to radioactivity decay, the parameter λ is called *decay constant*. In terms of the probabilistic interpretation, the waiting time T has a cumulative probability distribution

$$\Pr\{T \le t\} = 1 - P(t) = 1 - e^{-\lambda t}, \tag{6.3}$$

and a probability density function

$$f_T(t) = \frac{d}{dt}\Pr\{T \le t\} = \lambda e^{-\lambda t}, \quad (t \ge 0) \tag{6.4}$$

It mean is

$$\langle T \rangle = \int_0^\infty t f_T(t) dt = \frac{1}{\lambda}. \tag{6.5}$$

The reciprocal of the decay constant is the mean waiting time. The variance of the

$$\begin{aligned} \mathrm{Var}\big[T\big] &= \Big\langle \big(T - \langle T\rangle\big)^2 \Big\rangle = \big\langle T^2 - 2T\langle T\rangle + \langle T\rangle^2 \big\rangle \\ &= \big\langle T^2 \big\rangle - \langle T\rangle^2 = \int_0^\infty t^2 f_T(t)dt - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2}. \end{aligned} \tag{6.6}$$

Note the mean of a sum is always equal to the sum of the means.

If ther is a block of radioactive material within which there are total N_0 nuclei and each nuclear decay is statistically identical and independent, with the *probability* of not decayed at time t being $e^{-\lambda t}$, then the *remaining un-decayed* material at time t will be $N_0 e^{-\lambda t}$.

What happens if the decays of different nuclei are not independent?

6.1 Poisson process

Now let us consider independent rare events can occur again and again in time. For examples: shoppers arriving at a store on a remote street. Let $P_n(t)$ be the number of such event occurred up to time t. Then

$$P_{n+1}(t+\Delta t) = P_{n+1}(t)P_0(\Delta t) + P_n(t)P_1(\Delta t) + P_{n-1}(t)P_2(\Delta t) + \cdots \quad (6.7)$$

in which $P_2(\Delta t) = o(\Delta t)$, i.e., the probability of two such events to occur in the very short Δt time is negligible. Since $P_1(\Delta t) = \lambda \Delta t$, and

$$P_0(\Delta t) = 1 - \lambda \Delta t + o(\Delta t),$$

we have

$$\frac{P_{n+1}(t+\Delta t) - P_n(t)}{\Delta t} = -\lambda \Big[P_{n+1}(t) - P_n(t) \Big]. \quad (6.8)$$

In the limit of $\Delta t \to 0$, we have

$$\frac{dP_{n+1}(t)}{dt} = -\lambda \Big[P_{n+1}(t) - P_n(t) \Big]. \quad (6.9)$$

This system of equations has to be solved one by one, starting with

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \ \Rightarrow \ P_0(t) = e^{-\lambda t}.$$

Then,

$$\frac{dP_1(t)}{dt} = -\lambda \Big[P_1(t) - P_0(t) \Big] \ \Rightarrow \ P_1(t) = \lambda t e^{-\lambda t}.$$

$$\frac{dP_2(t)}{dt} = -\lambda \Big[P_2(t) - P_1(t) \Big] \ \Rightarrow \ P_2(t) = \frac{(\lambda t)^2}{2} e^{-\lambda t}.$$

$$\cdots$$

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \quad (6.10)$$

This is known as Poisson process. It is the *probability of having n events occurred at time t*. If we denote $N(t)$ as the number of events occurred at time t, then

$$\Pr \{ N(t) = n \} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \quad (6.11)$$

Interestingly, one can describe this process differently, in terms of the times at which each events occurs: $T_1, T_2, \cdots, T_n$, etc. Then

$$\Pr \Big\{ T_n > t \Big\} = \Pr \Big\{ N(t) \le n-1 \Big\} = \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \quad (6.12)$$

$$\Pr \Big\{ T_n \le t \Big\} = 1 - \Pr \Big\{ T_n > t \Big\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

$F_{T_n}(t) = \Pr\left\{T_n \le t\right\}$ is called a *cumulative probability distribution*. The corresponding probability density function is

$$\begin{aligned} f_{T_n}(t) &= \frac{dF_{T_n}(t)}{dt} = -\sum_{k=0}^{n-1} \frac{\lambda^k k t^{k-1}}{k!} e^{-\lambda t} + \sum_{k=0}^{n-1} \frac{\lambda^{k+1} t^k}{k!} e^{-\lambda t} \\ &= -\sum_{h=0}^{n-2} \frac{\lambda^{h+1} t^h}{h!} e^{-\lambda t} + \sum_{k=0}^{n-1} \frac{\lambda^{k+1} t^k}{k!} e^{-\lambda t} = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \quad (6.13) \end{aligned}$$

This is known as Gamma distribution of order n. It is the probability density function for the T_n, the random time at which the n^{th} event occurs. The first event occurs, T_1, has a distribution $\lambda e^{-\lambda t}$.

The distribution for T_n in Eq. 6.13 has another surprising interpretation. We note that

$$f_{T_2}(t) = \lambda^2 t e^{-\lambda t} = \int_0^t \left(\lambda e^{-\lambda s}\right)\left(\lambda e^{-\lambda(t-s)}\right) ds \qquad (6.14)$$

which is actually the probability distribution of the sum of two independent exponentially distributed $T_1^{(1)} + T_1^{(2)} = T_2$. In fact, it is easy to verify that

$$\frac{\lambda^{n+1} t^n}{n!} e^{-\lambda s} = \int_0^t \left(\frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda s}\right)\left(\lambda e^{-\lambda(t-s)}\right) ds. \qquad (6.15)$$

That is,

$$f_{T_n}(t) = \int_0^t f_{T_{n-1}}(s) f_{T_1}(t-s) ds, \text{ or } T_n = T_{n-1} + T_1. \qquad (6.16)$$

where T_n, T_{n-1}, and T_1 are three random variables.

6.2 Distribution for the time of the first event

If there are N identical, independently distributed (i.i.d.) random times $T_1, T_2, \cdots, T_N$ are all possible to occur, each follows an exponential waiting time with rate λ, then the first event to occur is

$$T^* = \min\left(T_1, T_2, \cdots, T_N\right). \qquad (6.17)$$

whose probability distribution is

$$\begin{aligned} \Pr\left\{T^* > t\right\} &= \Pr\left\{T_1 > t, T_2 > t, \cdots, T_N > t\right\} \\ &= \Pr\left\{T_1 > t\right\} \times \Pr\left\{T_2 > t\right\} \times \cdots \Pr\left\{T_N > t\right\}, \end{aligned}$$

in which

$$\Pr\left\{T_k > t\right\} = \int_t^\infty \lambda e^{-\lambda s} ds = e^{-\lambda t}.$$

$$\Pr\left\{T^* > t\right\} = \left(e^{-\lambda t}\right)^N = e^{-N\lambda t},$$

and it is again an exponentially distributed, but with the rate is N times faster.

Now, if we apply this result to the case of N boys and M girls at a party, and each boy encountering a girl with an exponential waiting time, as such events are always too rare(!), then the i^{th} encounters the j^{th} girl has the time T_{ij}. Now time for one of any boy to encounter with any girl is again exponential with rate $NM\lambda$. This is fundamental reason for we always using the "product term" to describe encountering between two individuals, be them people at a party, animals in an ecological system, or chemical species in a reaction.

6.3 Exponential waiting time is memoryless, memoryless has to be exponential

We now demonstrate a defining characteristic of exponential waiting time: *memoryless*. To understand this, we first has to learn the notion of *conditional probability*.

We denote the probability of an event A under the condition of another event B as $\Pr\{A|B\}$. For example, for the random variable T, the event A is $T \geq t_1$, and B is $T > t_2$. Then $\Pr\{T > t_1|T > t_2\}$ is the conditional probability of $T > t_1$ given $T > t_2$. Clear, if $t_1 \leq t_2$, then this conditional probability is 1; $T > t_1$ has to be true for sure if it is given that $T > t_2$. For $t_1 > t_2$, the conditional probabilities

$$\Pr\{T > t_1|T > t_2\} = \frac{\Pr\{T > t_1\}}{\Pr\{T > t_2\}}, \tag{6.18a}$$

and

$$\Pr\{T \leq t_1|T > t_2\} = \frac{\Pr\{t_2 < T \leq t_1\}}{\Pr\{T > t_2\}}. \tag{6.18b}$$

A random time T is memoryless if

$$\frac{\Pr\{T > t+s\}}{\Pr\{T > s\}} = \Pr\{T > t\}. \tag{6.19}$$

This means how long one waits for the event $T > t + s$ to occur is actually independent of when he or she starts to observe!

Eq. 6.19 implies that, if we denote the $\Pr\{T > t\} = g(t)$,

$$\frac{g(t+s)}{g(s)} = g(t).$$

Therefore,

$$g'(t+s) = g'(t)g(s) + g(t)g'(s)$$

which implies

$$\frac{g'(t)}{g(t)} = \frac{g'(s)}{g(s)} = -\lambda.$$

Therefore,

$$g(t) = Ae^{-\lambda t}.$$

Since $g(0) = 1$, $A = 1$. Thus T has to be exponentially distributed.

Conversely, it is easy to show that an exponentially distributed random time is memoryless.

6.4 Why is exponentially distributed time for an event so universal?

We have used the idea of "independent rare event" to derive this interesting distribution. We now give three examples:

(a) Changing light bulbs, superposition of i.i.d. renewal processes;

(b) The very first event among a large collection of i.i.d., non-exponential waiting times;

(c) Time for a barrier crossing — this is the foundation of radioactive decay and molecular isomerization reactions.

We shall demonstrate (b). The proof is actually quite similar to that in Sec. 6.2. Let $T_1, T_2, \cdots, T_N$ be N identical, independently distributed (i.i.d.) random times, each with the same probability density function (pdf) $f_T(t)$ and cumulative distribution function (cdf) $F_T(t)$. Then the first event to occur is

$$T^* = \min\big(T_1, T_2, \cdots, T_N\big). \tag{6.20}$$

whose probability distribution is

$$\begin{aligned}\Pr\Big\{T^* > t\Big\} &= \Pr\Big\{T_1 > t, T_2 > t, \cdots, T_N > t\Big\} \\ &= \Pr\Big\{T_1 > t\Big\} \times \Pr\Big\{T_2 > t\Big\} \times \cdots \Pr\Big\{T_N > t\Big\} \\ &= \Big(1 - F_T(t)\Big)^N.\end{aligned}$$

It is easy to see that when $N \to \infty$, $T^* \to 0$. Let us denote $\hat{T}^* = NT^*$. Then

$$\Pr\Big\{\hat{T}^* > s\Big\} = \Pr\Big\{T^* > s/N\Big\} = \left[1 - F_T\left(\frac{s}{N}\right)\right]^N.$$

If we let $N \to \infty$, then

$$\left[1 - F_T\left(\frac{s}{N}\right)\right]^N = \left[1 - F_T(0) - F_T'(0)\left(\frac{s}{N}\right) + o\left(N^{-1}\right)\right]^N,$$

$$\Pr\left\{\hat{T}^* > s\right\} \to e^{-\lambda s},$$

which is an exponentially distributed, with the rate $\lambda = F_T'(0) = f_T(0)$.

Therefore, with a large population of i.i.d. individuals, any event that occurs in the population, one by one, is almost always exponentially distributed in time: birth, death, catching a cold, getting married, etc.

6.5 von Neumann's rejection sampling

How do we get a "random number" that follows a given $f_X(x)$? Usually on the computer, one has a subroutine that gives a uniform distribution on $[0, 1]$. That is, the random variable with a pdf

$$f_{\text{uniform}}(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & x < 0, x > 1 \end{cases} \tag{6.21}$$

But what if you want to sample a distribution $f_X(x)$?

The rejection method works as follows. Let us assume $a \le X \le b$, and its pdf $0 \le f_X(x) \le c$, $x \in [a, b]$. Call the uniform random variable twice and obtain r_1 and r_2. Let

$$R_1 = a + (b - a)r_1, \;\; R_2 = cr_2.$$

Then accept the value R_1 if $R_2 \le f_X(R_1)$, and reject it and re-start if $R_2 > f_X(R_1)$. The R_1 you obtained in such a procedue follows the given $f_X(x)$.

If you are interested in sampling an exponentially distribution non-negative random variable T, $f_T(x) = \lambda e^{-\lambda x}$, then there is a more straightforward way:

$$T = -\frac{1}{\lambda} \ln r, \tag{6.22}$$

where r is uniformly distributed on $[0, 1]$.

To prove this is true:

$$\begin{aligned} F_T(x) &= \Pr\left\{0 \le T \le x\right\} = \Pr\left\{0 \le -\lambda^{-1} \ln r \le x\right\} \\ &= \Pr\left\{1 \ge r \ge e^{-\lambda x}\right\} = 1 - e^{-\lambda x}. \end{aligned}$$

This yields the pdf of T as

$$f_T(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_T(x) = \lambda e^{-\lambda x}.$$