Cool probabilistic things we typically don't teach our students about radioactive decay, but should

David Jon Furbish

Department of Earth and Environmental Sciences, Vanderbilt University, Nashville, Tennessee, USA

1 Initial remarks

In all respects the discovery of radioactive decay by Henri Becquerel in 1896, followed by the discovery of radium and polonium by Marie and Pierre Curie, profoundly influenced the thinking of atomic physicists in the early 1900s. Recognition of the implications, then applications, of radioactivity quickly followed, including the first use of radioactive decay by Bertram Boltwood in 1907 to date rock materials. Now, because numerous analytical techniques and analyses of systems hinge on the systematics of radioactive decay, descriptions of these systematics appear as a foundational topic in many of our Earth science courses. But here's the thing. Judging from texts and websites, we typically leave out important parts — the best parts — in our introductory descriptions. Specifically, we leave out key probabilistic elements of the radioactive decay process when we present the so-called equations of radioactive decay. As a consequence we risk offering misconceptions about what these equations actually represent, including the assumptions and uncertainty involved in using them. This as a lost opportunity for our students.

The purpose of this essay is to provide an introductory description of radioactive decay as viewed from a probabilistic perspective. In fact, an understanding of how radioactive decay works requires a probabilistic description; no other possibility exists. This starts with a description of radioactive decay as a Poisson process and then focuses on the mathematics of the systematics of decay rather than the nuclear physics of decay. The main text presents things mostly in a qualitative manner. For interested readers, the appendixes contain formal derivations of the Poisson and Binomial distributions for this problem. These appendixes also elaborate the idea of a *Gibbs ensemble* and the meaning of a *master equation*.

2 A typical presentation

We typically start by describing a time series of random decay events — a homogeneous Poisson process — emphasizing the agelessness of radioactive atoms, where the occurrence of a decay event is entirely independent of the occurrence of previous events. We define the decay constant λ as the expected number of decay events per unit time obtained as an average over an imagined long period of time. Then, if we have confidence in our students' abilities to appreciate differential equations, we say that the rate at which the number of atoms $N(t)$ changes with time t due to decay is proportional to the number present, and write

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\lambda N(t)\,. \qquad (1)$$

We then solve (1) to give (Appendix A)

$$N(t) = N_0 e^{-\lambda t}\,, \qquad (2)$$

where N_0 denotes the initial number of atoms at time $t = 0$ and the decay constant λ is specific to the radioactive element. With these results in place we might then move to, for example, a description of the relation between the decay constant and the half-life, decay chains, the idea of secular equilibrium between primary and product elements, radiometric dating methods, and so on.

This seems well and good. But here are things this introductory description leaves out. First, if (2) is interpreted at face value as the number of atoms remaining at time t, then this is wrong. In fact (1) and (2) are deterministic descriptions of a process that is purely probabilistic. The giveaway is that (2) admits real numbers for $N(t)$ when in fact this number must be an integer. What we typically fail to tell our students is that $N(t)$ in (1) and (2) represents the *expected* number of atoms at time t in relation to an *ensemble* of systems, not the actual number — a random variable — that occurs in any realization, except possibly by chance. Second, whereas we correctly appeal to the idea of a homogeneous Poisson process to describe how radioactive decay randomly occurs with rate constant λ, (1) and (2) actually represent the outcome of an *inhomogeneous* Poisson process. Radioactive decay starting with a finite number of atoms, as in (2), is only approximately a homogeneous Poisson process in the limit of large N over time scales much smaller than the half-life of the radioactive element. Third, that the correct description of the expected number of atoms given by (2) obtains from the differential equation (1) is for probabilistic reasons that are not physically represented by the deterministic form of (1).

Oh my. That's a lot of stuff associated with just two basic equations. So let's back up and fill in some of the pieces.

3 Foundational elements

3.1 Idealized radioactive decay as a homogeneous Poisson process

We say that radioactive decay is a Poisson process.[1] A Poisson process is a random process in which events occur randomly in time or space. The occurrence of an event is entirely independent of the occurrence of previous events. In addition to radioactive decay, examples of a Poisson process over time include the arrival of photons on a surface, the occurrence of raindrop impacts on a surface, and the events defined by bed load sediment particles crossing a line, for example the end of a flume. Poisson processes over space include the random placement of objects on a map, including the locations of photons and raindrop impacts on a surface.

Focusing on time, the probability that an event will occur within a specified interval is independent of the past, no matter how long ago a previous event occurred. If a long interval without an event occurs, the probability that an event will occur within the next specified interval does not increase; it remains unchanged. The number of events within a specified interval is entirely independent of the number of events in any other non-overlapping interval.

Let us imagine a great number of radioactive atoms. We observe them one at a time, and at the moment we start observing an atom we set time $t = 0$. We wait until each atom decays and record the time $t = w$ of decay. Alternatively we observe all of the atoms at once starting at time $t = 0$, then record the set of wait times w to decay. We will show below that the continuous probability distribution $f_w(w)$ [T^{-1}] of the wait times w [T] is exponential, namely,

$$f_w(w) = \frac{1}{\mu_w} e^{-w/\mu_w} \qquad w \geq 0\,, \tag{3}$$

where μ_w is the mean wait time. The variance of this distribution is $\sigma_w^2 = \mu_w^2$. For reference below the cumulative distribution function is $F_x(x) = 1 - e^{-w/\mu_w}$ so the exceedance probability function $R_w(w) = 1 - F_w(w) = e^{-w/\mu_w}$.

Let us now randomly sample a great number of wait times w from the distribution (3) and string them together to make a time series of decay events (Figure 1). Because the wait times are randomly selected, the occurrence of any decay event is entirely independent of the occurrence of previous events. This time series represents a homogeneous Poisson process with rate constant $\lambda = 1/\mu_w$ [T^{-1}]. That is, because μ_w is the average time between decay events, $\lambda = 1/\mu_w$ is the *expected* number of events per unit time. For reference below let us emphasize that the wait times

[1] Typically our descriptions of the randomness of radioactive decay imply a homogeneous Poisson process.

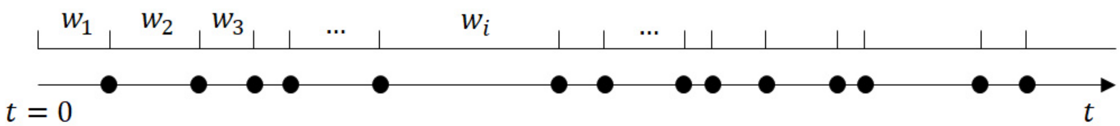

Figure 1: Schematic diagram of homogeneous Poisson events (circles) on a timeline created from randomly selected wait times $w_1, w_2, w_3, \ldots$ separating the events.

w are independent and identically distributed according to (3).

Let us now again randomly sample a great number of wait times from the distribution (3) and create a second time series of decay events. Due to the randomness of our sampling the occurrence of decay events in this second series is distinct from the events in the first series. In fact we can randomly sample wait times from the distribution (3) to create an arbitrarily large number of times series, each series being distinct from all others. What we have created is an *ensemble* of individual realizations (Appendix B) of a homogeneous Poisson process.

Now let n denote the total number of decay events at time t, and consider a plot of our results (Figure 2). In this example we are using a

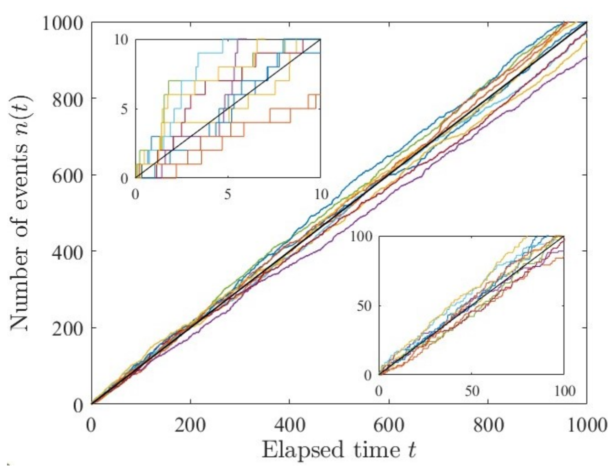

Figure 2: Plot of 10 realizations (colored lines) of the number of events $n(t)$ versus time t for a homogeneous Poisson process with rate constant $\lambda = 1$. Black line is the expected value λt.

relatively small rate constant $\lambda = 1$ in order to readily see the effects of the randomness in decay events. Notice that at any time t the number of events n varies among the individual realizations. That is, even though each realization is generated by the same algorithm using the same rate constant λ, by chance the number n varies. Let $p_n(n, t)$ denote the discrete distribution of the number of events n at time t. This distribution is given by

$$p_n(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad n \geq 0 \,, \quad t \geq 0 \,, \quad (4)$$

which is a Poisson distribution (Appendix C). The mean of this distribution is $\mu_n = \lambda t$ and its variance is $\sigma_n^2 = \lambda t$. That is, the variance is equal to the mean, and both increase linearly with time t. For an individual system there is one possible outcome n at time t with probability given by (4). For an ensemble of systems (realizations) all possible outcomes $n = 0, 1, 2, 3, \ldots$ exist at time t in the proportions given by (4).

Suppose we want to estimate the rate constant λ from an individual realization $n(t)$. This estimate of λ is $\hat{\lambda} = n(t)/t$. Because the number $n(t)$ differs among realizations, the estimate $\hat{\lambda}$ differs. The "true" value of λ is obtained from an individual realization only in the limit of $t \to \infty$. Because we can observe decay events only over a finite interval of time, we must accept uncertainty in our estimates $\hat{\lambda}$ of λ. This is reflected in our reporting of the uncertainty of the decay constants or half-lives of radioactive elements (e.g. Lucas and Unterweger, 2000; Pommé, 2015).

3.2 Exponential distribution of wait times

Let us now show that the distribution $f_w(w)$ of wait times w is exponential. Note that although we started with the exponential distribution (3) of wait times to create realizations of a Poisson process, the derivation of the Poisson distribution (Appendix C) given by (4) comes first.[2] Based on (4) the probability $p_n(0,t)$ that $n = 0$ events occur by time t is

$$p_n(0,t) = e^{-\lambda t}\,. \qquad (5)$$

This is the same as saying that $p_n(0,t)$ is the probability that we must wait to time $t = w$ to observe a decay event. Or, this is the probability that the wait time w exceeds the value $w = t$. If we reinterpret t as the wait time w, then

$$p_n(0,t) = R_w(w) = e^{-w/\mu_w} \qquad (6)$$

This is just the exceedance probability function of the wait times with $\lambda = 1/\mu_w$. It follows that the distribution $f_w(w)$ of wait times w is exponential and given by (3).

3.3 Radioactive decay as an inhomogeneous Poisson process

We now turn to radioactive decay as described by (1) and (2). First note that in order to describe radioactive decay as a homogeneous Poisson process as above we need to imagine a single atom, which, when the atom decays it is instantly replaced with another atom, and then another and so on to create a time series of decay events. Or, we might imagine a fixed number of atoms N such that each time an atom decays it is instantly replaced with another atom. But in this case the fixed, expected rate of decay is no longer equal to λ. Rather, the homogeneous Poisson rate is now equal to $\gamma = \lambda N$. The rate λ is the expected number of decay events per unit time per atom present; the rate $\gamma = \lambda N$ is the expected number of events per unit time for fixed N.

With a fixed number N of atoms involving replacement we can randomly sample a set of wait times w from an exponential distribution with mean $\mu_w = 1/\gamma$ to create a time series of decay events representing a homogeneous Poisson process. But if the number $N(t)$ changes with time t as in (1) and (2), the wait times between successive events are no longer independent and identically distributed. At any instant the Poisson rate is $\gamma = \lambda N(t)$. This represents an inhomogeneous Poisson process.

Consider a set of realizations of the number of atoms $N(t)$, each starting with the initial number N_0 (Figure 3). In this example we are using

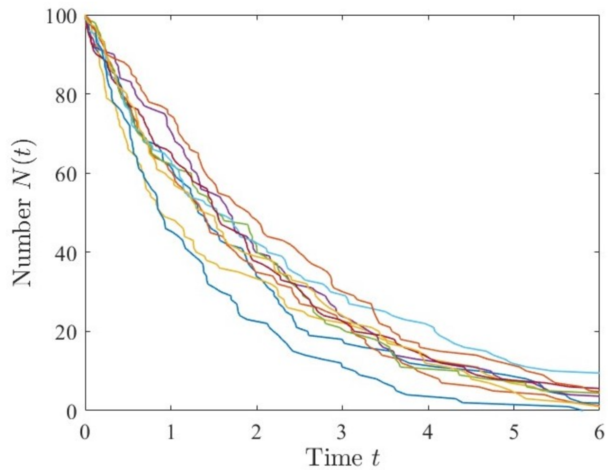

Figure 3: Plot of 10 realizations of the number of atoms $N(t)$ versus time t, each starting with $N_0 = 100$. In this example $\lambda = 1/2$.

a relatively small value of $N_0 = 100$ with $\lambda = 1/2$ in order to readily see the effects of the randomness in decay events. Notice that, like the realizations of a homogeneous Poisson process, the number $N(t)$ at any time t varies among the individual realizations. Let $p_N(N,t)$ denote the discrete distribution of the number N at time t. This distribution is given by

$$p_N(N,t) = \frac{N_0!}{N!(N_0-N)!}(e^{-\lambda t})^N(1-e^{-\lambda t})^{N_0-N}$$
$$0 \le N \le N_0\,. \qquad (7)$$

This is a binomial distribution with expected value $\mu_N = N_0 e^{-\lambda t}$, which is the result given by (2). For an individual system there is one possible outcome N at time t with probability given by (7). For an ensemble of systems all possible outcomes $N = 0, 1, 2, 3, \ldots$ exist at time t in the proportions given by (7) (Figure 4). Now imagine rotating this figure 90 degrees counterclockwise. One can then visualize that the expected value μ_N associated with the set of distributions

[2] The exponential distribution of wait times also can be independently deduced from survival theory.

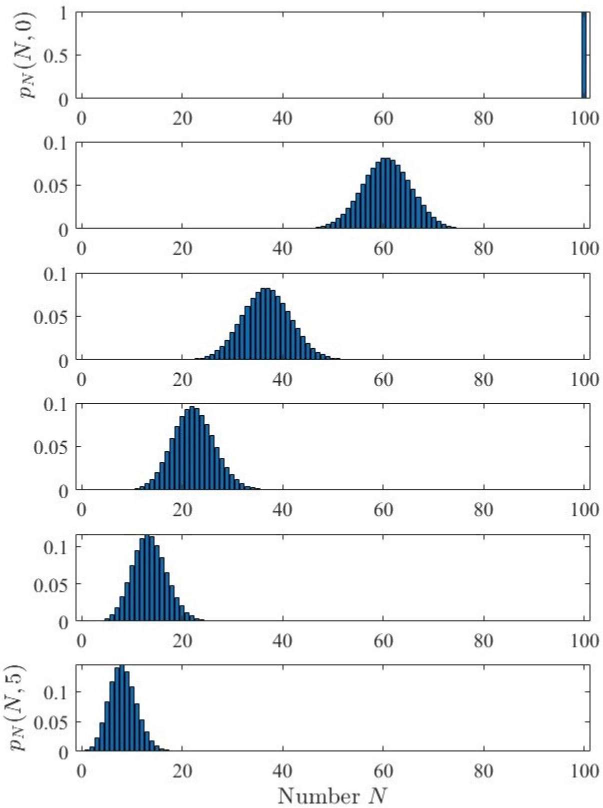

Figure 4: Plot of the distribution $p_N(N,t)$ at successive times $t = 0, 1, 2, 3, 4, 5$. These plots coincide with conditions represented by the realizations in Figure 3.

$p_N(N,t)$ declines exponentially. Letting angle brackets denote an ensemble expected (average) value, the correct way to write (1) and (2) is

$$\frac{\mathrm{d}\langle N(t)\rangle}{\mathrm{d}t} = -\lambda \langle N(t)\rangle \,, \tag{8}$$

and

$$\langle N(t)\rangle = N_0 e^{-\lambda t} \,. \tag{9}$$

The expected value $\langle N(t)\rangle$ is an imaginary thing associated with an ensemble; it is not something that happens in reality. As an expected value it may be a real number; it does not need to be an integer.

Because the number of decay events $n = N_0 - N$, one can show (Appendix D) that n is described by a binomial distribution. This also means that the number n represents the number of product atoms, neglecting decay of these atoms. The expected value $\langle n\rangle = N_0 - \langle N\rangle = N_0(1 - e^{-\lambda t})$.

3.4 The large–number–small–time limit

Let $t_\mathrm{m} = t_2 - t_1$ with $t_1 \geq 0$ denote a measurement interval. For sufficiently large N and $t_\mathrm{m} \ll 1/\lambda$ the binomial distribution is well approximated by a Poisson distribution. This means that over a sufficiently short interval of time t_m the relative change in the number N is small. The Poisson rate is $\gamma = \lambda N$ and the wait times w are well approximated by an exponential distribution with mean $\mu_w = 1/\gamma$. This is the basis of using Poisson statistics in evaluating uncertainty in measurements of the half-life, radiometric dating, and so on. It also is the basis for introducing the idea that radioactive decay involving a large number of atoms proceeds as a Poisson process.

4 Postscript

Radioactive decay is a rich topic whose implications and applications appear in many fields of science. Moreover, because of its familiarity, radioactive decay is a nice entry into the broader topic of stochastic processes. The idea of a Poisson process in particularly is a lovely starting point for considering a variety of stochastic processes that occur in natural and engineered systems across many scales, including compound Poisson processes, Lévy processes, renewal processes, OrnsteinUhlenbeck processes and so on.[3] Thus radioactive decay offers a nice initial glimpse into possibilities for describing noise driven systems, biotic as well as abiotic.

5 Appendix A: Obtaining the exponential solution for $N(t)$

Before solving (1) formally to give (2), let us do some math without actually doing any math. Namely, (1) tells us that the rate at which the

[3] Such processes are particularly relevant in my own work on the statistical physics of sediment transport.

number of atoms changes with time due to radioactive decay is proportional to the number present, and we want to know the function that expresses the number $N(t)$ in terms of time t. So, what function $N(t)$ do we know, which, when we take its first derivative with respect to time (the left side of (1)) is equal to itself multiplied by a constant (the right side of (1))? The answer, of course, is an exponential function. Thus, we know the form of the solution of (1) — an exponential function — before doing any formal math. This is a good thing, because we have a target in mind.

Now let us do things formally. We know the derivative of the function $N(t)$, so to obtain this function we must reverse the derivative operation, namely, take the anti-derivative. We start by separating the variables to give

$$\frac{1}{N}\,\mathrm{d}N = -\lambda\,\mathrm{d}t\,. \qquad (10)$$

That is, we place the dependent variable N and its differential $\mathrm{d}N$ on the left side, and we place the independent time differential $\mathrm{d}t$ on the right side. Taking the anti-derivative,

$$\int \frac{1}{N}\,\mathrm{d}N = -\lambda \int \mathrm{d}t\,. \qquad (11)$$

Evaluating the integrals then gives

$$\ln N = -\lambda t + C_1\,, \qquad (12)$$

where C_1 denotes the constant of integration. Exponentiation of both sides of (12) then gives

$$e^{\ln N} = e^{-\lambda t + C_1}\,. \qquad (13)$$

This means that

$$N(t) = e^{C_1} e^{-\lambda t} = C_2 e^{-\lambda t}\,, \qquad (14)$$

with the constant $C_2 = e^{C_1}$. We now note that the initial number of atoms $N(0) = N_0$ at time $t = 0$ so that $C_2 = N_0$. Substituting this result into (14) then gives (2) in the main text.

Appendix B: Ensembles

As used in this essay, the idea of an ensemble is attributable to Josiah W. Gibbs (Gibbs, 1902). Gibbs was focused on the statistical physics of gas particle systems.[4] Nonetheless, the idea of a Gibbs ensemble as currently used in statistical physics — a rather broad field — now is applied to many types of systems.

An ensemble is a great number, a set, of nominally identical but independent systems. It may be a real set, or, more usually, an imagined set of systems. In the case of radioactive decay we might imagine a great number of individual atoms, each observed starting at time $t = 0$. Or, we might imagine a great number of samples, each containing many radioactive atoms, where each sample starts with precisely N_0 atoms at time $t = 0$.

The first ensemble described above allows us to imagine observing the wait times to decay $w = t$ as described in Section 3.1, leading to the exponential distribution (3) then the Poisson distribution (4). The second ensemble described above allows us to imagine a great number of realizations of the number of atoms $N(t)$ during decay, leading to the binomial distribution (7). Similarly, in the following two appendixes an ensemble refers to a great number of realizations, each satisfying the same probabilistic rules.

Appendix C: Poisson distribution

Consider a timeline t over which events randomly occur with a fixed expected rate λ $[\mathrm{T}^{-1}]$ (Figure 5). This expected rate is just the average number of events per unit time. We now choose a small interval T. Because non-overlapping intervals of length T are independent, we can position this interval at many locations on the timeline and count the number of events within it. Notice that the number of events varies with the location. If we increase the interval T, on average it contains a larger number of events, but the number of events again varies with location. Moreover, we can imagine doing this for a great

[4] Gibbs coined the name *statistical mechanics* in his studies aimed at providing a theoretical explanation of the laws of thermodynamics. The field of statistical mechanics gave rise to the broader field of statistical physics.

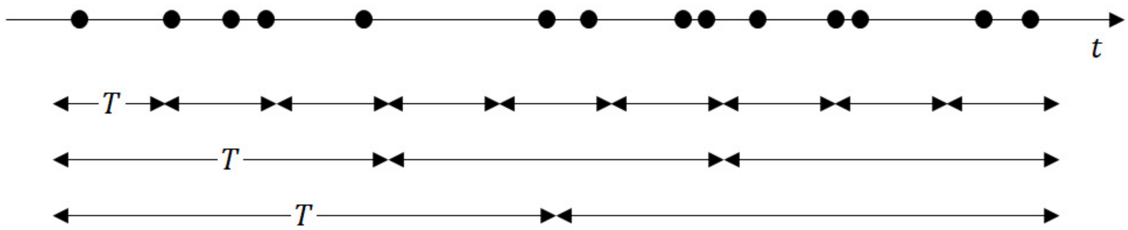

Figure 5: Schematic diagram of Poisson events (circles) randomly occurring on a timeline t, with intervals of varying length T containing different numbers of events.

number of intervals of varying length T. On average the number of events contained within each interval length T increases with this length.

Now choose a time axis during which events randomly occur for $t > 0$. We assume that the probability that an event will occur within any small interval $\mathrm{d}t$ is $\lambda\mathrm{d}t$ for rate constant λ. (We elaborate this idea below.) Let q denote the probability that no event will occur within $\mathrm{d}t$. Then assume that the probability that more than one event will occur during $\mathrm{d}t$ is smaller by an order of magnitude than the occurrence of one event during $\mathrm{d}t$, and denote this small probability as $O(\mathrm{d}t)$. These three possibilities are the only things that can occur during $\mathrm{d}t$, so $\lambda\mathrm{d}t + q + O(\mathrm{d}t) = 1$. This means that the probability q that no event will occur is

$$q = 1 - \lambda\mathrm{d}t + O(\mathrm{d}t)\,, \tag{15}$$

where we are unconcerned with the sign of $O(\mathrm{d}t)$. Finally, we assume that the number of events in non-overlapping time intervals are independent.

Following Feller (1949), let $p_n(n,t)$ denote the probability of having n events within the time interval $(0,t)$, and let $p_n(n,t+\mathrm{d}t)$ denote the probability of having n events within the interval $(0,t+\mathrm{d}t)$. We start by calculating the latter. There are three ways to reach a state of n events at time $t+\mathrm{d}t$. First, n events occur during the interval $(0,t)$ and no event occurs during $(t,t+\mathrm{d}t)$. Second, $n-1$ events occur during $(0,t)$ and one event occurs during $(t,t+\mathrm{d}t)$. Third, fewer than $n-1$ events occur during $(0,t)$ and more than one occurs during $(t,t+\mathrm{d}t)$. The probability $p_n(n,t+\mathrm{d}t)$ must equal the sum of the possible ways to reach this state, so for $n \geq 1$,

$$p_n(n,t+\mathrm{d}t) = [1 - \lambda\mathrm{d}t + O(\mathrm{d}t)]p_n(n,t)$$
$$+\lambda\mathrm{d}t p_n(n-1,t) + \mathrm{O}(\mathrm{d}t)\,. \tag{16}$$

Notice that the product $O(\mathrm{d}t)p_n(n,t)$ cannot be larger than $O(\mathrm{d}t)$, so we simplify this to

$$p_n(n,t+\mathrm{d}t) = (1 - \lambda\mathrm{d}t)p_n(n,t)$$
$$+\lambda\mathrm{d}t p_n(n-1,t) + O(\mathrm{d}t)\,. \tag{17}$$

The first term on the right side of (17) is the probability that n events occur during $(0,t)$ and no event occurs during $(t,t+\mathrm{d}t)$. The second term on the right side is the probability that $n-1$ events occur during $(0,t)$ and one occurs during $(t,t+\mathrm{d}t)$. The last term represents the occurrence of fewer than $n-1$ events preceding t and multiple events during $(t,t+\mathrm{d}t)$, and essentially represents the error of the approximation given by the first two terms. For the state $p_n(0,t+\mathrm{d}t)$,

$$p_n(0,t+\mathrm{d}t) = [1 - \lambda\mathrm{d}t + O(\mathrm{d}t)]p_n(0,t)\,, \tag{18}$$

which we may simplify to

$$p_n(0,t+\mathrm{d}t) = (1 - \lambda\mathrm{d}t)p_n(0,t) + O(\mathrm{d}t)\,. \tag{19}$$

The first term on the right side of (19) represents the probability that no event occurs during $(0,t)$ and no event occurs during $(t,t+\mathrm{d}t)$. The last term represents the error of the approximation.

We now expand (17) and (19), rearrange, then divide by $\mathrm{d}t$ to give

$$\frac{p_n(n,t+\mathrm{d}t) - p_n(n,t)}{\mathrm{d}t} = -\lambda p_n(n,t)$$
$$+\lambda p_n(n-1,t) + \frac{O(\mathrm{d}t)}{\mathrm{d}t} \quad \text{and} \tag{20}$$
$$\frac{p_n(0,t+\mathrm{d}t) - p_n(0,t)}{\mathrm{d}t} = -\lambda p_n(0,t)$$

$$+\frac{O(\mathrm{d}t)}{\mathrm{d}t}\,. \tag{21}$$

Taking the limit as $\mathrm{d}t \to 0$ then gives

$$\frac{\partial p_n(n,t)}{\partial t} = -\lambda p_n(n,t)$$

$$+\lambda p_n(n-1,t) \qquad \text{and} \tag{22}$$

$$\frac{\partial p_n(0,t)}{\partial t} = -\lambda p_n(0,t)\,, \tag{23}$$

where we are assuming that in this limit, $O(\mathrm{d}t)/\mathrm{d}t \to 0$. The initial conditions for (22) and (23) are

$$p_n(0,0) = 1 \qquad \text{and} \qquad p_n(n,0) = 0\,. \tag{24}$$

Together, (22) and (23) describe the rate at which the probability $p_n(n,t)$ for $n = 0, 1, 2, 3, ...$ changes as the interval t increases. Note that partial derivatives (22) and (23) are used because the probability distribution $p_n(n,t)$ is a function of both time t and the number state n. We may interpret $p_n(n,t)$ as the proportion of an ensemble of systems (Appendix B) that are in the state n at time t, in which case (22) and (23) are considered master equations.

According to (22),

$$\frac{\partial p_n(1,t)}{\partial t} = -\lambda p_n(1,t) + \lambda p_n(0,t)\,, \tag{25}$$

$$\frac{\partial p_n(2,t)}{\partial t} = -\lambda p_n(2,t) + \lambda p_n(1,t) \tag{26}$$

and so on for $n = 3, 4, 5, \ldots$. Starting with (23) and using (24) the solution is

$$p_n(0,t) = e^{-\lambda t}\,. \tag{27}$$

Substituting this result into (25) and using (24) the solution is

$$p_n(1,t) = \lambda t e^{-\lambda t}\,. \tag{28}$$

In turn, substituting this result into (26) and using (24) the solution is

$$p_n(2,t) = \frac{1}{2}\lambda^2 t^2 e^{-\lambda t}\,. \tag{29}$$

Continuing this recursive operation then leads by inspection to the general result that

$$p_n(n,t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad n \geq 0, \quad t \geq 0\,. \tag{30}$$

This is the Poisson distribution with mean $\mu_n = \lambda t$ and variance $\sigma_n^2 = \lambda t$. This distribution is unusual in that its variance is equal to the mean. As $\mu_n = \lambda t$ increases this distribution increasingly has the appearance of a Gaussian distribution. Indeed, the Gaussian distribution is a good approximation of the Poisson distribution for a sufficiently large value of μ_n. For an individual system there is one possible outcome n at time t with probability given by (30). For an ensemble of systems all possible outcomes $n = 0, 1, 2, 3, \ldots$ exist at time t in the proportions given by (30).

Let us return to the idea above that $\lambda \mathrm{d}t$ may be considered a probability. This product is $\mathrm{d}t/\mu_w$. Now, if λ is the expected number of events per time and $\mu_w = 1/\lambda$ is the expected time per event, the ratio $\mathrm{d}t/\mu_w$ is the proportion of this expected time per event associated with $\mathrm{d}t$. Thus $\lambda \mathrm{d}t$ may be interpreted as the probability that an event will occur during $\mathrm{d}t$.

Appendix D: Binomial distribution

The derivation of the binomial distribution for the problem of radioactive decay is similar to the derivation of the Poisson distribution in Appendix C. We start by letting λ denote the probability per unit time that a decay event will occur per number of radioactive atoms present. Then $\lambda \mathrm{d}t$ is the probability that an event will occur during any small interval $\mathrm{d}t$ per number of radioactive atoms present, and $\lambda N \mathrm{d}t$ is the probability that an event will occur during $\mathrm{d}t$. Let q denote the probability that no decay event will occur during $\mathrm{d}t$. Then assume that the probability that more than one event will occur during $\mathrm{d}t$ is smaller by at least an order of magnitude than the occurrence of one event during $\mathrm{d}t$, and denote this small probability as $O(\mathrm{d}t)$, to mean on the order of the interval $\mathrm{d}t$. These three possibilities are the only things that can occur during $\mathrm{d}t$, so $\lambda N \mathrm{d}t + q + O(\mathrm{d}t) = 1$. This means that the probability q that no event will occur is

$$q = 1 - \lambda N \mathrm{d}t + O(\mathrm{d}t)\,, \tag{31}$$

where we are unconcerned with the sign of $O(\mathrm{d}t)$. Finally, we assume that the number of events in non-overlapping time intervals are independent.

Let $p_N(N,t)$ denote the probability that a system has N radioactive atoms at time t and let $p_N(N,t+\mathrm{d}t)$ denote the probability that the system has N radioactive atoms at time $t+\mathrm{d}t$. There are two ways to reach this latter state, so

$$p_N(N,t+\mathrm{d}t) = p_N(N+1,t)(N+1)\lambda\mathrm{d}t$$
$$+p_N(N,t)(1-N\lambda\mathrm{d}t) + O(\mathrm{d}t)\,. \qquad (32)$$

The first term on right side of (32) is the probability that a system has $N+1$ atoms at time t and an atom is lost by decay during $\mathrm{d}t$. The second term on the right side is the probability that a system has N atoms at time t and no atom is lost. The last term represents the error of the approximation provided by the two preceding terms. For an initial state defined by $N=N_0$,

$$p_N(N_0,t+\mathrm{d}t) = p_N(N_0,t)(1-N_0\lambda\mathrm{d}t)$$
$$+O(\mathrm{d}t)\,. \qquad (33)$$

The first term on the right side of (33) is the probability that a system has N_0 atoms at time t and no atom is lost. The last term represents the error of the approximation.

We now expand (32) and (33), rearrange, then divide by $\mathrm{d}t$ and take the limit as $\mathrm{d}t \to 0$ to give

$$\frac{\partial p_N(N,t)}{\partial t} = \lambda(N+1)p_N(N+1,t)$$
$$-\lambda N p_N(N,t) \qquad \text{and} \qquad (34)$$
$$\frac{\partial p_N(N_0,t)}{\partial t} = -\lambda N_0 p_N(N_0,t)\,. \qquad (35)$$

The initial conditions are

$$p_N(N_0,0) = 1 \qquad \text{and}$$
$$np_N(N,0) = 0\,, \quad N < N_0\,. \qquad (36)$$

Note that partial derivatives (34) and (35) are used because the probability distribution $p_N(N,t)$ is a function of both time t and the number state N. Further note that (34) and (35) are referred to as master equations in that they describe the time evolution of the distribution $p_N(N,t)$ for all possible states N. We now obtain this distribution using a recursive solution as follows.

Separating the variables in (35) and integrating gives

$$\ln[p_N(N_0,t)] = -\lambda N_0 t + C_1\,. \qquad (37)$$

Exponentiation then gives

$$p_N(N_0,t) = C_2 e^{-\lambda N_0 t}\,, \qquad (38)$$

with constant $C_2 = e^{C_1}$. Using the initial condition in (36), that $p_N(N_0,0)=1$ at time $t=0$, then $C_2=1$ and

$$p_N(N_0,t) = e^{-\lambda N_0 t}\,, \qquad (39)$$

which shows that the probability $p_N(N_0,t)$ decreases exponentially. That is, at time $t=0$ the probability is one that the system has N_0 radioactive atoms. Then, the probability that the system retains N_0 radioactive atoms without any decay decreases exponentially with time. Now notice that, according to (34), for $N=N_0-1$,

$$\frac{\partial p_N(N_0-1,t)}{\partial t} = \lambda N_0 p_N(N_0,t)$$
$$-\lambda(N_0-1)p_N(N_0-1,t)\,. \qquad (40)$$

Substituting (39) then gives

$$\frac{\partial p_N(N_0-1,t)}{\partial t} = \lambda N_0 e^{-\lambda N_0 t}$$
$$-\lambda(N_0-1)p_N(N_0-1,t)\,. \qquad (41)$$

With a bit of work we can integrate this equation to give

$$p_N(N_0-1,t) = C_1 e^{-\lambda(N_0-1)t} - N_0 e^{-\lambda N_0 t}\,. \quad (42)$$

Using the initial condition that $p_N(N_0-1,0)=0$ at time $t=0$, then $C_1=N_0$ and

$$p_N(N_0-1,t) = N_0 e^{-\lambda N_0 t}(e^{\lambda t}-1)\,. \qquad (43)$$

Proceeding recursively we eventually conclude by inspection that the distribution of the number of radioactive atoms at time t is

$$p_N(N,t) = \frac{N_0!}{N!(N_0-N)!}(e^{-\lambda t})^N(1-e^{-\lambda t})^{N_0-N}$$

$$0 \leq N \leq N_0\,. \tag{44}$$

This is a binomial distribution with expected value $N_0 e^{-\lambda t}$, which is the result given by (2). For an individual system there is one possible outcome N at time t with probability given by (44). For an ensemble of systems all possible outcomes $N = 0, 1, 2, 3, \ldots$ exist at time t in the proportions given by (44).

The number of decay events $n = N_0 - N$. In this situation substitution into (44) leads to the result that n also is described by a binomial distribution,

$$p_n(n,t) = \frac{N_0!}{n!(N_0-n)!}(e^{-\lambda t})^{(N_0-n)}(1-e^{-\lambda t})^n$$

$$0 \leq n \leq N_0\,. \tag{45}$$

The expected value $\langle n \rangle = N_0 - \langle N \rangle = N_0(1 - e^{-\lambda t})$ (Figure 6).

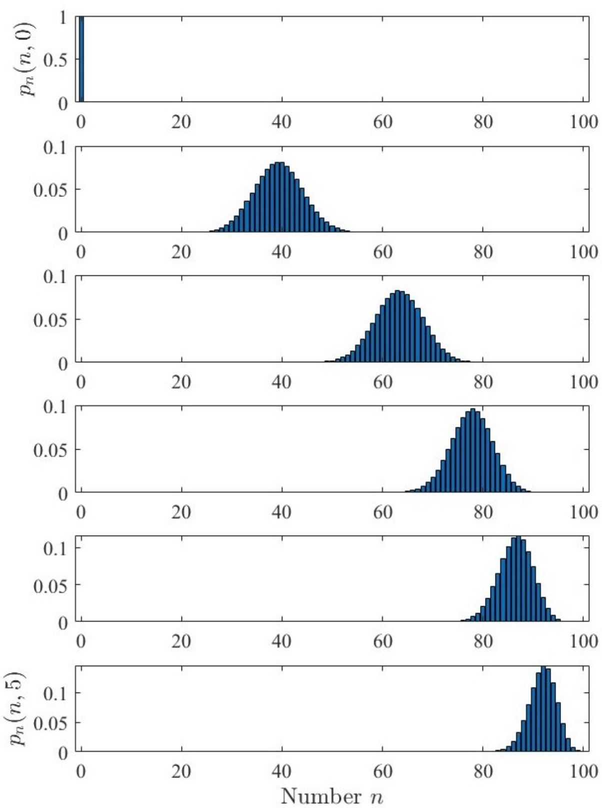

Figure 6: Plot of the distribution $p_n(n,t)$ at successive times $t = 0, 1, 2, 3, 4, 5$. These plots coincide with conditions represented in Figure 3 and Figure 4.

References

[1] Feller, W. (1949) On the theory of stochastic processes, with particular reference to applications, *Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, California, 403–432, https://projecteuclid.org/euclid.bsmsp/1166219215.

[2] Gibbs, J. W. (1902) *Elementary Principles in Statistical Mechanics*, Yale University Press, New Haven, Connecticut.

[3] Lucas, L. L. and Unterweger, M. P. (2000) Comprehensive review and critical evaluation of the half-life of tritium, *Journal of Research of the National Institute of Standards and Technology*, 105, 541–549.

[4] Pommé, S. (2015) The uncertainty of the half-life, *Metrologia*, 52, S51–S65.