

Chapter 6

Exponential Distribution and the Process of Radioactive Decay

The problem of radioactive decay is one of the simplest examples which show the connection between deterministic models and stochastic approach. We emphasize that the stochastic model is not an alternative to the deterministic approach, rather it is a generalization: *a satisfactory stochastic model gives a derivation for the deterministic model and more.*

The deterministic model for radioactive decay $A \rightarrow A^*$ is a simple linear differential equation

$$\frac{dN_A}{dt} = -kN_A$$

where N_A is the number of atom A and k is the rate of decay. To solve this ODE with initial condition $N_A(0) = N_0$, we have

$$N_A(t) = N_0 e^{-\lambda t}. \quad (6.1)$$

If we observe the decay of atom A one by one, then one realizes that the time at which A transforming to A^* is random. Hence the microscopic process of radioactive decay has to be modeled by a stochastic model. Specifically, if we denote \mathbf{X} as the lifetime of A , \mathbf{X} is a continuous, positive random variable. How do we determine the probability distribution for \mathbf{X} ? The basic assumption for the stochastic model is that the decay of A , as a random event, is independent of the time the atom being in A . In other words, the conditional probability

$$Prob\{\mathbf{X} > t + h | \mathbf{X} > t\} = Prob\{\mathbf{X} > h\}.$$

As we will soon see, this assumption is sufficient to determine the probability distribution of \mathbf{X} . It is known as the exponential distribution.

6.1 Exponential Distribution

The exponential distribution has the very important property known as memoriless:

$$Prob\{\mathbf{X} > t + h | \mathbf{X} > t\} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = Prob\{\mathbf{X} > h\}.$$

In fact, this nice property is the defining character of the exponential distribution. Consider a RV \mathbf{X} with

$$\begin{aligned} Prob\{\mathbf{X} > t + h\} &= Prob\{\mathbf{X} > t + h | \mathbf{X} > t\} Prob\{\mathbf{X} > t\} \\ &= Prob\{\mathbf{X} > h\} Prob\{\mathbf{X} > t\}. \end{aligned}$$

Let

$$G(t) = Prob\{\mathbf{X} > t\},$$

then

$$\begin{aligned} G(t + h) &= G(t)G(h) \\ G'(t + h) &= G'(t)G(h) = G(t)G'(h) \\ \implies \frac{G'(t)}{G(t)} &= \frac{G'(h)}{G(h)} = -\lambda \\ \implies G'(t) &= -\lambda G(t) \end{aligned}$$

i.e., \mathbf{X} is exponential

$$Prob\{\mathbf{X} \leq t\} = 1 - G(t) = 1 - e^{-\lambda t}. \quad (6.2)$$

6.2 Microscopic versus Macroscopic Models

What is the relation between Eqn. 6.1 and Eqn. 6.2. After all, both are mathematical models for the radioactive decay. The relation lies upon the a system of N number of independent atoms with very large N . From Eqn. 6.2, we note that at any time t , the atom is still in A with probability $e^{-\lambda t}$ and in A^* with probability $1 - e^{-\lambda t}$. This is a binary distribution. Hence, for N_0 iid, we have

$$Prob\{\mathbf{N}_t = n\} = \frac{N_0!}{n!(N_0 - n)!} e^{-n\lambda t} (1 - e^{-\lambda t})^{N_0 - n} \quad (6.3)$$

where RV \mathbf{N}_t is the number of atoms being A at time t . So the expectation and variance for \mathbf{N}_t are

$$E[\mathbf{N}_t] = N_0 e^{-\lambda t}, \quad Var[\mathbf{N}_t] = N_0 e^{-\lambda t} (1 - e^{-\lambda t})$$

which indicates that Eqn. 6.1 is simply the mean of the \mathbf{N}_t . The stochastic model, however, also provide an estimation for the variance. It is shown that for a large N_0 on the order of 10^{20} , the relative broadness of the distribution

$$\frac{Var[\mathbf{N}_t]}{E[\mathbf{N}_t]^2} \propto \frac{1}{N_0} \approx 0.$$

This proves that the deterministic model is a very good approximation if one deals with large number of atoms, i.e., macroscopic. It also shows when studying system of only a few number of atoms (microscopic), the stochastic model is more general.

Chapter 7

Poisson Processes

A Poisson process is associated with the repeatedly happening of an event along a one-dimensional axis (called time). These events can be represented by points randomly distributed on the time-axis.

- (i) all events are independent in each disjoint interval;
- (ii) the probability of having number of events in a very small interval $(t, t + \tau)$ is

$$P_1(\tau) = \lambda\tau + o(\tau), \quad \text{and} \quad P_0(\tau) = 1 - P_1(\tau)$$

where λ is called the intensity of the Poisson Process. Then we have:

$$P_k(\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}.$$

Proof 3 To show this, we note that

$$P_0(t + dt) = P_0(t)P_0(dt) = P_0(t) [1 - \lambda dt + o(dt)]$$

$$\implies P_0'(t) = -\lambda P_0(t)$$

and since $P_0(0) = 1$, we have

$$P_0(t) = e^{-\lambda t}.$$

For $k \geq 1$,

$$P_k(t+dt) = P_k(t)P_0(dt) + P_{k-1}(t)P_1(dt) + o(dt) = P_k(t)(1 - \lambda dt) + P_{k-1}(t)(\lambda dt) + o(dt),$$

$$\frac{P_k(t + dt) - P_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t) + o(1),$$

$$\frac{dP_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t). \quad (7.1)$$

To solve Eqn. 7.1, we introduce $\psi_k(t) = P_k(t)e^{\lambda t}$, and equation for $\psi_k(t)$:

$$\begin{aligned}
\frac{d\psi_k(t)}{dt} &= \frac{dP_k(t)}{dt} e^{\lambda t} + \lambda\psi_k(t) \\
&= -\lambda\psi_k(t) + \lambda\psi_{k-1}(t) + \lambda\psi_k(t) \\
&= \lambda\psi_{k-1}(t)
\end{aligned}$$

with condition $\psi_k(0) = 0$. We therefore have

$$\psi_k(t) = \frac{(\lambda t)^k}{k!}, \quad \text{and} \quad P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda}.$$

7.1 Properties of a Poisson Process

There are two aspects of a Poisson process: the distribution for the time intervals and the distribution for the corresponding **counting process**.

(i) if $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ are independent Poisson random variables, then

$$\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2 + \dots + \mathbf{Z}_m$$

is still Poisson.

Homework. Try to use the method of generating function to show that the random sum of

$$\mathbf{N}_1 + \mathbf{N}_2 + \dots + \mathbf{N}_K,$$

where \mathbf{K} is a Poisson RV with mean λ , and \mathbf{N} 's are iid binary RVs with $Prob\{\mathbf{N} = 1\} = p$ and $Prob\{\mathbf{N} = 0\} = 1 - p$, is still Poisson.

(ii) The waiting distribution. What is the probability distribution of first arriving event? Let's denote the time by a RV \mathbf{T}_1 , then

$$\begin{aligned}
Prob\{\mathbf{T}_1 \leq t\} &= 1 - Prob\{\mathbf{T}_1 > t\} \\
&= 1 - Prob\{\mathbf{N}_t = 0\} \\
&= 1 - e^{-\lambda t}
\end{aligned}$$

where \mathbf{N}_t is the number of event before time t (counting process). The pdf for \mathbf{T}_1 is obtained by derivative:

$$f_{\mathbf{T}_1}(t) = \lambda e^{-\lambda t}.$$

Hence, Poisson process is intimately related to the exponential distribution.

Similarly, let's consider \mathbf{T}_k :

$$\begin{aligned}
Prob\{\mathbf{T}_k \leq t\} &= 1 - Prob\{\mathbf{T}_k > t\} \\
&= 1 - Prob\{\mathbf{N}_t = k - 1\} \\
&= 1 - \sum_{\ell=0}^{k-1} \frac{(\lambda t)^\ell e^{-\lambda t}}{\ell!}.
\end{aligned}$$

Again, differentiation with respect to t , we have pdf:

$$\begin{aligned} f_{\mathbf{T}_k}(t) &= \frac{d}{dt} \text{Prob}\{\mathbf{T}_k \leq t\} = - \sum_{\ell=0}^{k-1} \frac{\lambda \ell (\lambda t)^{\ell-1} - \lambda (\lambda t)^\ell e^{-\lambda t}}{\ell!} \\ &= - \sum_{\ell=0}^{k-2} \frac{\lambda (\lambda t)^\ell e^{-\lambda t}}{\ell!} + \sum_{\ell=0}^{k-1} \frac{\lambda (\lambda t)^\ell e^{-\lambda t}}{\ell!} \\ &= \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!} \end{aligned}$$

which is called *gamma distribution*. When $k = 1$, \mathbf{T}_k is called exponentially distributed.

(iii) We now show that gamma distribution of order k is just the sum of k exponential distributions:

$$\mathbf{T}_k = \mathbf{T}_1 + (\mathbf{T}_2 - \mathbf{T}_1) + (\mathbf{T}_3 - \mathbf{T}_2) + \dots + (\mathbf{T}_k - \mathbf{T}_{k-1}),$$

where the waiting times for i th event $\mathbf{T}_i - \mathbf{T}_{i-1}$ are independent.

Proof 4 We first find out the characteristic function for gamma distribution:

$$\begin{aligned} \int_0^\infty \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!} e^{-ist} dt &= - \left(\frac{\lambda}{\lambda + is} \right) \int_0^\infty \frac{(\lambda t)^{k-1}}{(k-1)!} d e^{-(\lambda + is)t} \\ &= \left(\frac{\lambda}{\lambda + is} \right) \int_0^\infty \frac{\lambda (\lambda t)^{k-2}}{(k-2)!} e^{-(\lambda + is)t} dt \\ &= \dots \\ &= \left(\frac{\lambda}{\lambda + is} \right)^k. \end{aligned}$$

Note that

$$\frac{\lambda}{\lambda + is} = \int_0^\infty \lambda e^{-\lambda t} e^{-ist} dt$$

is the characteristic function of an exponential distribution.

(iv) The mean of the gamma distribution is

$$\left. \frac{\partial}{\partial s} \left(\frac{\lambda}{\lambda + is} \right)^k \right|_{s=0} = - \left. \frac{ik\lambda^k}{(\lambda + is)^{k+1}} \right|_{s=0} = - \frac{ik\lambda^k}{\lambda^{k+1}} = -i \frac{k}{\lambda},$$

as expected. $1/\lambda$ is the mean waiting time.

7.2 Uniform Distribution and Poisson Processes

If points generated by a Poisson process are labeled on a time axis, what is their distribution? The answer is they are uniform with density λ . This is not a very intuitive

result. The key is that the statement is *conditioned on a fixed total number of events in an interval* !

Let $\mathbf{T}_1, \mathbf{T}_2, \dots$ be the occurrence times in a Poisson process of intensity λ . Conditioned on $\mathbf{N}_t = n$, the random variable $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_n$ are uniformly distributed on the interval $[0, t)$.

We prove a simplified version of this result: $n = 1$. We know joint pdf for RVs \mathbf{T}_1 and \mathbf{T}_2 :

$$Prob\{t \leq \mathbf{T}_1 < t + dt, \tau \leq \mathbf{T}_2 < \tau + d\tau\} = \lambda^2 e^{-\lambda t} e^{-\lambda(\tau-t)} dt d\tau \quad (\tau > t),$$

therefore,

$$Prob\{t \leq \mathbf{T}_1 < t + dt, \mathbf{T}_2 \geq \tau\} = \int_{\tau}^{\infty} \lambda^2 e^{-\lambda t} e^{-\lambda(\tau'-t)} dt d\tau' = \lambda (1 - e^{-\lambda \tau}) dt, \quad (\tau > t).$$

Hence,

$$\begin{aligned} Prob\{t \leq \mathbf{T}_1 < t + dt | \mathbf{T}_2 \geq \tau\} &= \frac{Prob\{t \leq \mathbf{T}_1 < t + dt, \mathbf{T}_2 \geq \tau\}}{Prob\{\mathbf{T}_2 \geq \tau\}} \quad (t < \tau) \\ &= \frac{\lambda (1 - e^{-\lambda \tau}) dt}{\int_0^{\tau} \lambda (1 - e^{-\lambda \tau}) dt} \quad (t < \tau) \\ &= \frac{dt}{\tau} \quad (t < \tau). \end{aligned}$$

This is a uniform distribution on $[0, \tau)$. Note that if without the condition,

$$Prob\{t \leq \mathbf{T}_1 < t + dt\} = \lambda e^{-\lambda t} dt \quad (0 \leq t < \infty).$$

This result shows how important the condition is to a probabilistic problem.

7.3 Poissonization

Poisson processes not only offer models for biological problems, they also provide an approach to some problems that at first sight seem totally unrelated. Formulating such problems in terms of Poisson processes is called Poissonization. The relationship between a Poisson process and the uniform distribution is the foundation for this approach.

Let's consider, for instance, computing the expectation $E(\mathbf{N}_{b,g})$ of the total number of children $\mathbf{N}_{b,g}$ born to a couple who stop reproducing when they reach their goal of at least b boys and at least g girls. For simplicity, we assume that each birth represents an independent trial with the two possible outcomes, boy, with probability p , and girl, with probability $q = 1 - p$.

How do Poisson processes possibly come into play? It is useful to view the births as spaced events in time according to the random times generated by a Poisson process of unit intensity ($\lambda = 1$). Thus, on average n births occur during $[0, n)$ for any positive integer n . When births are classified by sex, the random number of boys born during

$[0, t)$ is independent of the random number of girls born during $[0, t)$ for every $t \geq 0$. Therefore, in essence there are two independent Poisson processes operating in parallel. Let \mathbf{T}_b and \mathbf{T}_g be the waiting time times until the birth of b boys and g girls, respectively. The waiting time until at least b boys and g girls arrive is

$$\mathbf{T}_{b,g} = \max(\mathbf{T}_b, \mathbf{T}_g).$$

Therefore,

$$\begin{aligned} E[\mathbf{T}_{b,g}] &= \int_0^\infty \text{Prob}\{\mathbf{T}_{b,g} > t\} dt \\ &= \int_0^\infty [1 - \text{Prob}\{\mathbf{T}_{b,g} \leq t\}] dt \\ &= \int_0^\infty [1 - \text{Prob}\{\mathbf{T}_b \leq t\} \text{Prob}\{\mathbf{T}_g \leq t\}] dt \\ &= \int_0^\infty \text{Prob}\{\mathbf{T}_b > t\} dt + \int_0^\infty \text{Prob}\{\mathbf{T}_g > t\} dt - \int_0^\infty \text{Prob}\{\mathbf{T}_b > t\} \text{Prob}\{\mathbf{T}_g > t\} dt, \end{aligned}$$

and \mathbf{T}_b and \mathbf{T}_g are gamma distributions. Hence

$$E[\mathbf{T}_{b,g}] = \frac{b}{p} + \frac{g}{q} - \sum_{k=0}^{b-1} \sum_{\ell=0}^{g-1} \frac{(k+\ell)!}{k! \ell!} p^k q^\ell. \quad (7.2)$$

We now show that this result is consistent with an alternative derivation. It is easy to show that the $E[\mathbf{N}_{b,g}]$ satisfies the following difference equation:

$$E[\mathbf{N}_{b,g}] = p(E[\mathbf{N}_{b-1,g}] + 1) + q(E[\mathbf{N}_{b,g-1}] + 1) = pE[\mathbf{N}_{b-1,g}] + qE[\mathbf{N}_{b,g-1}] + 1,$$

with boundary conditions

$$\begin{aligned} E[\mathbf{N}_{0,g}] &= E[\mathbf{N}_{0,g-1}] + q + 2pq + 3p^2q + \dots \\ &= E[\mathbf{N}_{0,g-1}] + \frac{1}{q} \\ &= g/q, \end{aligned}$$

and similarly,

$$E[\mathbf{N}_{b,0}] = b/p.$$

We now show that Eqn. 7.2 is the solution. We first note that for any b and g ,

$$\begin{aligned}
& \sum_{k=0}^b \frac{(g+k)!}{k!g!} p^k q^{g+1} + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g \left(\frac{(g+k)(1-p)}{g} \right) + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g + \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g \left(\frac{k-p(g+k)}{g} \right) + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g + \sum_{k=1}^b \frac{(g+k-1)!}{(k-1)!g!} p^k q^g - \sum_{k=0}^b \frac{(g+k)!}{k!g!} p^{k+1} q^g + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g + \sum_{k=0}^{b-1} \frac{(g+k)!}{k!g!} p^{k+1} q^g - \sum_{k=0}^b \frac{(g+k)!}{k!g!} p^{k+1} q^g + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g - \frac{(b+g)!}{b!g!} p^{b+1} q^g + \sum_{\ell=0}^g \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \\
= & \sum_{k=0}^b \frac{(g+k-1)!}{k!(g-1)!} p^k q^g + \sum_{\ell=0}^{g-1} \frac{(b+\ell)!}{\ell!b!} p^{b+1} q^\ell \quad (\text{i.e., we have reduced } g \text{ by } 1) \\
= & \dots = \sum_{k=0}^{b-1} p^k q + p^b = 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& pE[\mathbf{T}_{b-1,g}] + qE[\mathbf{T}_{b,g-1}] + 1 \\
= & b-1 + \frac{pg}{q} - p \sum_{k=0}^{b-2} \sum_{\ell=0}^{g-1} \frac{(k+\ell)!}{k!\ell!} p^k q^\ell + \frac{qb}{p} + g-1 - q \sum_{k=0}^{b-1} \sum_{\ell=0}^{g-2} \frac{(k+\ell)!}{k!\ell!} p^k q^\ell + 1 \\
= & \frac{g}{q} + \frac{b}{p} - 1 + p \sum_{\ell=0}^{g-1} \frac{(b+\ell-1)!}{(b-1)!\ell!} p^{b-1} q^\ell + q \sum_{k=0}^{b-1} \frac{(k+g-1)!}{k!(g-1)!} p^k q^{g-1} - \sum_{k=0}^{b-1} \sum_{\ell=0}^{g-1} \frac{(k+\ell)!}{k!\ell!} p^k q^\ell \\
= & \frac{g}{q} + \frac{b}{p} - \sum_{k=0}^{b-1} \sum_{\ell=0}^{g-1} \frac{(k+\ell)!}{k!\ell!} p^k q^\ell \\
= & E[\mathbf{T}_{b,g}]
\end{aligned}$$

7.4 Radioactive Decay of Few Atoms and Bulk Material

We have shown that the number of decay events up to time t for total N_0 atoms at $t = 0$ is

$$Prob\{N_t^* = n\} = \frac{N_0!}{n!(N_0 - n)!} e^{-(N_0 - n)\lambda t} (1 - e^{-\lambda t})^n.$$

On the other hand, it is well known that the arriving of a X-ray particle due to radioactive decay is a Poisson process. That means the number of decay event up to time t is

$$Prob\{N_t^* = n\} = \frac{(\mu t)^n}{n!} e^{-\mu t}.$$

The relationship between these two models is that the latter is a bulk material with $N_0 \approx \infty$ but $N_0 \lambda = \mu$. In fact, it can be shown that in the limit of $N_0 \rightarrow \infty$

$$\begin{aligned} \frac{N_0!}{n!(N_0 - n)!} e^{-(N_0 - n)\lambda t} (1 - e^{-\lambda t})^n &\rightarrow \frac{N_0^n}{n!} e^{-N_0 \lambda t} (\lambda t)^n \\ &= \frac{(\mu t)^n}{n!} e^{-\mu t}. \end{aligned}$$

This reveals that the intensity μ for the Poisson process is in fact proportional to the total radioactive material. It is important to realize that for two independent exponentially distributed RVs \mathbf{X}_1 and \mathbf{X}_2 , the distribution for $\mathbf{Y} = \min\{\mathbf{X}_1, \mathbf{X}_2\}$ is

$$Prob\{\mathbf{Y} > y\} = Prob\{\mathbf{X}_1 > y\} Prob\{\mathbf{X}_2 > y\} = e^{-2\lambda y}.$$

Hence, the pdf for \mathbf{Y}

$$f_Y(y) = 2\lambda e^{-2\lambda y}$$

which has twice the rate of \mathbf{X} 's.

Chapter 8

Markov Chains

8.1 A Genetic Model (S. Wright)

We deal with fixed population of $2N$ genes composed of type a and A individuals. Let \mathbf{X}_n denote the number of type a gene in the n th generation. If $\mathbf{X}_n = j$, that is

$$a : j \qquad A : 2N - j,$$

then the make-up of the the next generation is determined by $2N$ independent binomial trials with each trial giving a with probability $p_j = j/2N$ and giving A with probability $q_j = 1 - j/2N$. The process of generation after generation is a Markov process with transition probability

$$Prob\{\mathbf{X}_{n+1} = k | \mathbf{X}_n = j\} = P_{jk} = \binom{2N}{k} p_j^k q_j^{2N-k}.$$

If there are spontaneous mutations with probability α for each $a \rightarrow A$ prior to the formation of the new generation, and probability β for each $A \rightarrow a$, then one has

$$p_j = \frac{j}{2N} - \alpha \frac{j}{2N} + \beta \left(1 - \frac{j}{2N}\right)$$

and

$$q_j = \left(1 - \frac{j}{2N}\right) + \alpha \frac{j}{2N} - \beta \left(1 - \frac{j}{2N}\right).$$

8.2 A Disease Spreading Model

A very simple model for the spread of a disease assumes a population of total N individuals, of which some are diseased and the remainder are healthy. During any single period of time, two people are selected at random from the population and assumed to have an encounter. If one of the two people is diseased and the other not, then with probability α (known as transmission coefficient) the disease is transmitted

to teh healthy person. Otherwise, no disease transmission takes place. Let \mathbf{X}_n be the number of diseased persons in the polulation at the end of n th period.

If $\mathbf{X}_n = j$, then at the end of $n + 1$ period, there is either still j or $j + 1$ diseased persons. The corresponding probabilities are:

$$P\{\mathbf{X}_{n+1} = j + 1 | \mathbf{X}_n = j\} = \alpha \frac{2j(N - j)}{N(N - 1)}$$

and

$$P\{\mathbf{X}_{n+1} = j | \mathbf{X}_n = j\} = 1 - \alpha \frac{2j(N - j)}{N(N - 1)}$$

where $N(N - 1)/2$ is the total number of possible number of pairing in the whole population, $j(N - j)$ is the possible number of paring between a diseased and a healthy individual, $j(j - 1)/2$ is the possible number of paring among the diseased population, and $(N - j)(N - j - 1)/2$ is the possilbe number of paring among the healthy individuals:

$$\frac{j(j - 1)}{2} + j(N - j) + \frac{(N - j)(N - j - 1)}{2} = \frac{N(N - 1)}{2}.$$

8.3 Markov Matrix and Stationary State

A **Markovian matrix** of order n is a $n \times n$ square matrix $P = [p_{ij}]$:

$$p_{ij} \geq 0, \quad \sum_{j=1}^n p_{ij} = 1.$$

It is easy to show that a Markov matrix has a right eigen vector $(1, 1, \dots, 1)$ with corresponding eigenvalue 1. Let's denote the corresponding left eigenvector $(\mu_1^*, \mu_2^*, \dots, \mu_n^*)$, then

$$\sum_{i=1}^n \mu_i^* p_{ij} = \mu_j^*.$$

$\{\mu^*\}$ is called stationary distribution.

For a Markov chain with discrete steps and states, the transition probability - the conditional probability from step n to $n + 1$ - is a Markov matrix. A markov chain with a constant transition matrix at each step is called homogeneous. It is easy to show that matrix P^2 is the transition probability from step n to $n + 2$, and matrix P^m is the transition probability from step n to $n + m$. In the limit of $m \rightarrow \infty$,

$$(P^m)_{ij} = \mu_j^*.$$

Hence the stationary state is also the limiting distribution of a Markov chain, which is independent of the initial distribution. To show this, note first that none of the eigenvalues of a Markov matrix is no greater than 1.

We now give two theorems which are sufficient and necessary for P_{ij} being Markovian.

Theorem 9 A diagonalizable matrix P_{ij} is Markovian if and only if:

$$P_{ij} = \sum_{k=0} \lambda_k a_i^k b_j^k$$

where:

$$\sum_{k=0} a_i^k b_j^k = \sum_k a_i^k b_k^j = \delta_{ij}$$

and:

$$\lambda_0 = 1, a_i^0 > 0, \sum_i a_i^0 = 1, b_j^0 = 1 \text{ for } \forall j, |\lambda_k| \leq 1, \sum_{k=0} \lambda_k \geq 0.$$

Proof 5 The proof for necessary condition can be in any text book on positive matrix. Now let's consider the sufficient condition:

$$P_{ij} = \sum_{k=0} \lambda_k a_i^k b_j^k = a_i^0 + \sum_{k=1} \lambda_k a_i^k b_j^k$$

hence if the second term is non-negative, $P_{ij} \geq 0$, if it is negative but $i \neq j$:

$$\begin{aligned} P_{ij} &= a_i^0 + \sum_{k=1} \lambda_k a_i^k b_j^k = a_i^0 - \left| \sum_{k=1} \lambda_k a_i^k b_j^k \right| \geq a_i^0 - \left| \sum_{k=1} |\lambda_k| a_i^k b_j^k \right| \\ &\geq a_i^0 - \left| \sum_{k=1} a_i^k b_j^k \right| = a_i^0 - \left| \sum_{k=0} a_i^k b_j^k - a_i^0 \right| = a_i^0 - | - a_i^0 | = 0 \end{aligned}$$

if $i = j$, then:

$$P_{ii} = a_i^0 - \left| \sum_{k=1} \lambda_k a_i^k b_i^k \right| \geq a_i^0 - \left| \sum_{k=1} \lambda_k \right| \left| \sum_{k=1} a_i^k b_i^k \right| \geq a_i^0 - \left| \sum_{k=0} a_i^k b_i^k - a_i^0 \right|$$

and:

$$\sum_i P_{ij} = 1 + \sum_i \sum_{k=1} \lambda_k a_i^k b_j^k = 1 + \sum_{k=1} \lambda_k \sum_i a_i^k b_j^k = 1 + \sum_{k=1} \lambda_k \sum_i a_i^k b_i^0 b_j^k = 1$$

8.4 Reversibility of a Markov Chain

Let's now consider a 2-state markov chain:

$$\begin{pmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{pmatrix}$$

Its stationary state is

$$(\mu_1^*, \mu_2^*) = \left(\frac{p_{21}}{p_{12} + p_{21}}, \frac{p_{12}}{p_{12} + p_{21}} \right).$$

That is

$$\mu_1^* p_{12} = \mu_2^* p_{21} \tag{8.1}$$

This result is similar to the equilibrium in chemical kinetics. The stationarity is sustained by a balance.

Will such a balance necessary for the stationary state of a Markov chain? Let's consider a 3-state Markov chain:

$$\begin{pmatrix} 1 - p_{12} - p_{13} & p_{12} & p_{13} \\ p_{21} & 1 - p_{21} - p_{23} & p_{23} \\ p_{31} & p_{32} & 1 - p_{31} - p_{32} \end{pmatrix}$$

Solving for the stationary state leads to

$$\mu_1^* p_{12} - \mu_2^* p_{21} = \mu_2^* p_{23} - \mu_3^* p_{32} = \mu_3^* p_{31} - \mu_1^* p_{13} \quad (8.2)$$

It is clear that Eqn. 8.1 implies that all terms in Eqn. 8.2 equal 0. Clearly, this is not a necessary condition for the stationarity. Such a situation is called *reversibility* or *detail balance*, two important concept in equilibrium thermodynamics.

Detail balance requires that

$$\frac{p_{12}p_{23}p_{31}}{p_{21}p_{32}p_{13}} = 1$$

or

$$\frac{p_{13}}{p_{31}} = \frac{p_{12}p_{23}}{p_{21}p_{32}} \quad (8.3)$$

These formula should be familiar to chemists and physicists. Eqn. 8.3 suggests that the reversible system has *path independent* properties. Hence a potential function can be introduced. The potential function is the well-known free energy.

For irreversible system, its stationary state is sustained by a nonzero flux. Such a flux generates entropy, and has to be “pumped” by external energy. This flux is known in Onsager-Hill's theory of irreversible thermodynamics.

8.5 Further Mathematics on Markov Matrix (Optional)

If all elements of a Markov matrix are positive, then the matrix is said **irreducible**.

Frobenius Theorem. If a matrix P is irreducible, then (a) it has an positive eigenvalue λ_0 with corresponding eigenvector also being positive (probability distribution). (b) $\lambda_0 = 1$. (c) λ_0 is the spectral radius of matrix P . (d) The eigenvector is unique.

(a) Let's consider all the real numbers λ to each of which corresponding a vector $x = (x_1, x_2, \dots, x_n)$

$$\sum_{i=1}^n x_i = 1, \quad x_i > 0, \quad \text{and} \quad xP \leq \lambda x \quad (8.4)$$

We define $\lambda_0 = \text{sup}\{\lambda\}$ is the lowest upper-bound of all λ .

One can show that $\lambda_0 \leq nM$ where M is the largest element of P . That is to say no x will give $xP \geq nMx$. This is because for any x ,

$$(xP)_i = \sum_{k=1}^n x_k P_{ki} < M \sum_{k=1}^n x_k = M.$$

for all the component i of the xP . On the other hand, if there was a y which satisfies $yP \geq nMy$. Then there would be at least one j so that $y_j \geq 1/n$. Hence

$$(yP)_j \geq nMy_j \geq M$$

This is a contradiction. Similarly, one can show that $\lambda_0 \geq nm$ where $m > 0$ is the smallest element of P .

We now show that λ_0 is an eigenvalue of P . By the definition of λ_0 , we have a sequence $\gamma_1, \gamma_2, \dots \rightarrow \lambda_0$ and vector x^1, x^2, \dots . Since this sequence of vectors are bound, i.e., their components lie in the interval $[0, 1]$, there is a subsequence which converge to x^* . Clearly, $x^*P \geq \lambda_0 x^*$. We now show that $x^*P = \lambda_0 x^*$, for otherwise, $x^*P > \lambda_0 x^*$. There we can find a $\lambda' > \lambda_0$ such that $x^*P > \lambda' x^*$. That contradicts λ_0 being the upper-bound of λ .

Therefore, λ_0 is positive and an eigenvalue of P . Its corresponding eigenvector is positive.

(b) Following (a), we have:

$$1 = \sum_{ij} x_i^* P_{ij} = \lambda_0 \sum_j x_j^* = \lambda_0$$

(c) Let $\lambda \neq \lambda_0$ be any other eigenvalue of P . We now show $|\lambda| < \lambda_0$. This is because

$$|\lambda||x| = |xP| \leq |x|P$$

where x is the corresponding eigenvector for λ . Hence, as the upper-bound, $\lambda_0 \geq |\lambda|$.

(d) Let's have another eigenvector $y \neq cx^*$, where c is any constant. Then we can find a μ such that $(x - \mu y)$ is still the eigenvector of λ_0 , $(x - \mu y) \geq 0$ but with at least one component being 0. This contradicts (a).

The probability distribution $\pi = x^*$ is called the **stationary distribution** of the Markov chain with transition probability P_{ij} :

$$\sum_{i=1}^n \pi_i P_{ij} = \pi_j$$

Chapter 9

Continuous Time Markov Chains

9.1 From Deterministic to Stochastic

Rate equations and mass-action. Let's consider the simplest chemical reaction:



where the forward and backforward rate constants are k_+ and k_- , respectively. A and B can be the two possible conformation of a six-carbon ring (cyclohexane) molecule which can be either a “boat” or a “chair” conformation. The freshman chemistry taught us that two rate equations can be set up:

$$\frac{d[A]}{dt} = -k_+[A] + k_-[B], \quad (9.1)$$

$$\frac{d[B]}{dt} = k_+[A] - k_-[B]. \quad (9.2)$$

One can solve this equation, given initial a condition, to learn all about the kinetics this simple chemical reaction.

However, if one looks only a few molecules, what do you expect? How about when only a single molecule is under observation? This simple question leads to the generalization of the rate equations above in terms of a Markov process. The above equations are based on the “law of mass action”. “mass” here means a large number of molecules.

The single molecule is constantly going back-and-forth between conformations A and B . When it is in A , its probability of going to B is like the radioactive decay with an exponential distribution, i.e., linear rate equation.

$$\frac{dP_A}{dt} = -k_+P_A, \quad \text{and} \quad \frac{dP_B}{dt} = k_+P_A.$$

Similarly, if the molecule is in B , then

$$\frac{dP_A}{dt} = k_- P_B, \quad \text{and} \quad \frac{dP_B}{dt} = -k_- P_B.$$

Therefore we have a set of equations:

$$\frac{dP_A}{dt} = -k_+ P_A + k_- P_B, \quad (9.3)$$

$$\frac{dP_B}{dt} = k_+ P_A - k_- P_B. \quad (9.4)$$

This is the model for unimolecular reaction in terms of a continuous time Markov chain.

Analysis of the stochastic unimolecular reaction. We now carry out an analysis for the stochastic model in terms of Eqs. 9.3 and 9.4. Clearly, $P_A + P_B = 1$. What is the interpretation for the stationary solution:

$$P_A = \frac{k_-}{k_+ + k_-}, \quad P_B = \frac{k_+}{k_+ + k_-}.$$

How is Eqs. 9.3 and 9.4 related to Eqs. 9.1 and 9.2?

The answer to the first question is that at stationary state, the molecule spends a fraction of P_A time in state A and a fraction of P_B time in state B . To show this, we note that the life-time of the molecule in state A (or B) is an exponential distribution with pdf $k_+ e^{-k_+ t}$ (or $k_- e^{-k_- t}$). Hence the mean life-time (also known as the sojourn time) is $1/k_+$ for state A and $1/k_-$ for state B .

To answer the second question, let's consider a system of N identical but independent molecules, each follows the stochastic model. This leads to a binomial distribution with expectations for RVs N_A and N_B satisfying the Eqs. 9.1 and 9.2.

Continuous-time and discrete-time markov chain. The general continuous-time Markov chain is a generalization of the Eqs. 9.3 and 9.4:

$$\begin{aligned} \frac{dP_1}{dt} &= -(\lambda_{12} + \lambda_{13} + \dots + \lambda_{1n})P_1 + \lambda_{21}P_2 + \lambda_{31}P_3 + \dots + \lambda_{n1}P_n \\ \frac{dP_2}{dt} &= \lambda_{12}P_1 - (\lambda_{21} + \lambda_{23} + \dots + \lambda_{2n})P_2 + \lambda_{21}P_3 + \dots + \lambda_{n2}P_n \\ &\dots \\ \frac{dP_n}{dt} &= \lambda_{1n}P_1 + \lambda_{2n}P_2 + \dots + \lambda_{n-1,n}P_{n-1} - (\lambda_{n1} + \lambda_{n2} + \dots + \lambda_{n,n-1})P_n \end{aligned} \quad (9.5)$$

This equation is sometime known as the master equation in physics. The $n \times n$ matrix $\Lambda = [\lambda_{ij}]$ has the following properties:

$$\lambda_{ij} \geq 0 \quad (i \neq j), \quad \sum_{j=1}^n \lambda_{ij} = 0.$$

It can be shown that all the eigenvalues are negative except one which is 0. We now introduce an exponential matrix representation for the continuous-time Markov chain.

If we choose Δt as a time step, then $P_{ij}(\Delta t)$ is a transition probability for a discrete-time Markov chain. Solving the linear ODE in Eqn. 9.6, we have

$$P_{ij}(\Delta t) = [e^{\Lambda \Delta t}]_{ij}. \quad (9.6)$$

The matrix Λ is called infinitesimal transition rate:

$$\Lambda = \lim_{\Delta \rightarrow 0} \frac{P_{ij}(\Delta t) - \delta_{ij}}{\Delta t}.$$

Note that $P_{ij}(0) = \delta_{ij}$.

9.2 Some Basic Equations

The Chapman-Kolmogorov Equation. Let $X(t)$ be a homogeneous, continuous time Markov chain on the states $s_1, s_2, \dots, s_n, \dots$. Suppose $P_{ij}(t)$ is the probability that the process is in state s_j at time t , given it starts in state s_i at time 0:

$$P_{ij}(t) = \text{Prob}\{X(t+s) = j | X(s) = i\}.$$

Then

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(s). \quad (9.7)$$

This equation states that in order to move from state s_i to state s_j in time $t+s$, $X(t)$ moves to some state s_k in time t and then from s_k to s_j in the remaining time s . This is the continuous-time analog of matrix multiplication for discrete-time Markov chains.

Equation (1) can be transformed into differential equations for $P_{ij}(t)$ if we know the infinitesimal transition rates λ_{ij} for the continuous-time Markov chains. For a very small t , $P_{ii}(t)$ represents approximately the probability that the process has *not* escaped from s_i . Hence

$$P_{ii}(t) = e^{-\lambda_i t} + o(t) = 1 - \lambda_i t + o(t),$$

where $\lambda_i = -\sum_{k \neq i} \lambda_{ik}$. Similarly,

$$P_{ij}(t) = \lambda_{ij} t + o(t)$$

Now using the Chapman-Kolmogorov equation (1) we have

$$\begin{aligned} P_{ij}(t+h) &= P_{ij}(t)P_{jj}(h) + \sum_{k \neq j} P_{ik}(t)P_{kj}(h) \\ &= P_{ij}(t)(1 - \lambda_j h) + \sum_{k \neq j} P_{ik}(t)\lambda_{kj} h + o(h) \end{aligned}$$

Therefore,

The Forward Equation.

$$\frac{dP_{ij}(t)}{dt} = -\lambda_j P_{ij}(t) + \sum_{k \neq j} \lambda_{kj} P_{ik}(t) \quad (9.8)$$

The forward equation are deduced by decomposing the time interval $(0, t + h)$, where h is positive and small, into two periods

$$(0, t), \quad (t, t + h)$$

and examing the transitions in each period separately. The initial condition for the differential equation is

$$P_{ij}(0) = \delta_{ij}$$

A different result arises from splitting the time interval $(0, t + h)$ into the two periods

$$(0, h), \quad (h, t + h)$$

and adapting the preceding analysis. We then have

The Backward Equation.

$$\frac{dP_{ij}(t)}{dt} = -\lambda_i P_{ij}(t) + \sum_{k \neq j} \lambda_{ik} P_{kj}(t). \quad (9.9)$$

The forward and backward equations can be best understood if we consider Eqn. 9.6. Differentiate it with respect to t :

$$\frac{dP_{ij}(t)}{dt} = \Lambda e^{\Lambda t} = e^{\Lambda t} \Lambda.$$

Hence, the forward and backward equations are commutation of the product of matrix Λ and P . Note that Eqn. 9.6 is a special case of the forward and backward equations which in general have time-dependent λ 's. The master equation characterizes a constant-rates continuous-time markov chain.

9.3 The Pure Birth Processes

A birth process is a Markov process $X(t)$ with a family of increasing staircase functions. The $X(t)$ takes value $1, 2, 3, \dots$ and increases by 1 at the discontinuous point t_i (birth time). The transition rates λ_{ij} are nonzero only if

$$\lambda_{i, i+1} = \mu_i.$$

All other non-diagonal λ_{ij} are zero, and the diagonal

$$-\lambda_i = \mu_i.$$

The pure birth process is a generalization of the Poisson process where $\mu_i \equiv \mu$.

Just as a Poisson process can be understood from either its counting process or the time-intervals, a pure birth process can also be understood by considering the random sojourn time \mathbf{S}_k . RV \mathbf{S}_k is the time between the k th and the $(k + 1)$ th birth events. Hence,

$$P_n(t) = Prob \left\{ \sum_{i=0}^{n-1} \mathbf{S}_i \leq t \leq \sum_{i=0}^n \mathbf{S}_i \right\}.$$

It is easy to show that, as in a Poisson process, all \mathbf{S}_k have exponential distribution with respective mean λ_k .

A particular birth process which characterizes the growth of a population with identical and independent individuals is the Yule process which has $\lambda_{n,n+1} = n\alpha$. The α is known as the growth rate *per capita* per time, the same α in the deterministic exponential growth model $dN(t)/dt = \alpha N(t)$.

9.4 The Birth&Death Processes

Similar to the birth processes, one can define a death process which has nonzero $\lambda_{i,i-1}$. The birth&death process is a generalization of the pure birth and pure death processes. It is the continuous-time analogue of a random walk with non-uniform bias.

The stationary solution to a birth&death processes can be explicitly obtained. We note that the stationary probability p_i^s satisfies

$$p_i^s \lambda_{i,i+1} = p_{i+1}^s \lambda_{i+1,i}. \quad (9.10)$$

Therefore, we have

$$p_i^s = C \prod_{\ell=1}^{i-1} \frac{\lambda_{\ell-1,\ell}}{\lambda_{\ell,\ell-1}}. \quad (9.11)$$

The C in the equation is a normalization factor which can be determined noting $\sum_{i=0}^{\infty} p_i^s = 1$.

One specific Birth&Death process is the Kendall's birth-death-immigration process in which $\lambda_{j,j-1} = j\mu$, representing the rate of death, and $\lambda_{j,j+1} = \nu + j\alpha$ representing the rate of birth ($\propto j\alpha$) and a constant rate of immigration (ν).

The equation for the probability distribution at time t follows the systems of equations

$$\begin{aligned} \frac{dp_0(t)}{dt} &= -\nu p_0(t) + \mu p_1(t) \\ \frac{dp_j(t)}{dt} &= -(\nu + j\alpha + j\mu)p_j(t) + (\nu + (j-1)\alpha)p_{j-1}(t) + (j+1)\mu p_{j+1}(t) \quad (j = 1, 2, \dots) \end{aligned}$$

To deal with the time-dependent solution to this set of equations, we define generating function

$$G(s, t) = \sum_{j=0}^{\infty} p_j(t) s^j, \quad (0 \leq s \leq 1) \quad (9.12)$$

and we have:

$$\boxed{\frac{\partial G(s, t)}{\partial t} = (\alpha s - \mu)(s - 1) \frac{\partial G(s, t)}{\partial s} + \nu(s - 1)G(s, t)}. \quad (9.13)$$

If the initial state of the Markov process is in state i , then $G(s, 0) = s^i$ corresponding to $p_j(0) = \delta_{ji}$. To obtain the moments for the birth-death-immigration process, the PDE can be simplified to ODEs. Take the mean as an example,

$$m = \left. \frac{\partial G(s, t)}{\partial s} \right|_{s=1}.$$

Hence, differentiating (9.12) with respect to s :

$$\frac{dm}{dt} = \frac{\partial^2}{\partial t \partial s} G(1, t) = (\alpha - \mu)m(t) + \nu$$

and $m(0) = i$. This ODE can be solved and we have

$$m(t) = i e^{-\int_0^t [\alpha(z) - \mu(z)] dz} + \int_0^t \nu(\tau) e^{\int_\tau^t [\alpha(z) - \mu(z)] dz} d\tau.$$

For the special case of constant, time-independent α , μ and ν , we have

$$m(t) = i e^{(\alpha - \mu)t} + \frac{\nu}{\alpha - \mu} (e^{(\alpha - \mu)t} - 1).$$

When $\alpha < \mu$, that is the birth rate is less than the death rate,

$$m(t) \longrightarrow \frac{\nu}{\mu - \alpha} \quad \text{as } t \longrightarrow \infty$$

Another example for the continuous-time Markov chain is the non-homogeneous Poisson process, which is a pure birth process with $\alpha(t) = \mu(t) = 0$, but ν is not a constant but a function of time, $\nu(t)$. One can show that the Eq. (9.13) is valid even when α , μ and ν are functions of time. Then for initial condition $i = 0$, we have the equation for a non-homogeneous Poisson process

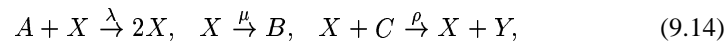
$$\frac{\partial}{\partial t} G(s, t) = \nu(t)(s - 1)G(s, t)$$

$$G(s, t) = e^{-(1-s) \int_0^t \nu(\tau) d\tau}$$

$$P_{0j}(t) = e^{-\int_0^t \nu(\tau) d\tau} \frac{\left(\int_0^t \nu(\tau) d\tau \right)^j}{j!}.$$

9.5 Birth& Death Process with Mutation

Interestingly, the population model for both linear birth and death, as well as mutation, can be represented by the simple chemical reaction. Be more specific, let X and Y are the numbers of individuals in two related populations, each with its own birth and death. However, there is a mutation from population X to Y . Now consider a chemical reaction system in which the number of molecules for species X and Y (by abusing the notation intentionally) change with time according to the reaction scheme



in which the parameters λ, μ, ρ are the birth rate, death rate, and mutation rate, respectively. Similarly, for the Y , we have



The λ', μ' might or might not be the same as the λ, μ .

