The Kemeny constant of a Markov chain

Peter Doyle

Version 1.0 dated 14 September 2009
GNU FDL[*]

Abstract

Given an ergodic finite-state Markov chain, let M_{iw} denote the mean time from i to equilibrium, meaning the expected time, starting from i, to arrive at a state selected randomly according to the equilibrium measure w of the chain. John Kemeny observed that M_{iw} does not depend on starting the point i. The common value $K = M_{iw}$ is the *Kemeny constant* or *seek time* of the chain. K is a spectral invariant, to wit, the trace of the resolvent matrix. We review basic facts about the seek time, and connect it to the bus paradox and the Central Limit Theorem for ergodic Markov chains.

For J. Laurie Snell

The seek time

We begin by reviewing basic facts and establishing notation for Markov chains. For background, see Kemeny and Snell [4] or Grinstead and Snell [3], bearing in mind that the notation here is somewhat different.

Let P be the transition matrix of an ergodic finite-state Markov chain. We write the entries of P using tensor notation, with $P_i{}^j$ being the probability that from state i we move to state j. (There is some possibility

---

here of confusing superscripted indices with exponents, but in practice it should be clear from context which is meant.) The sequence of matrix powers $I, P, P^2, P^3, \ldots$ has a (Cesaro) limit which we will denote by P^∞. We have $PP^\infty = P^\infty P = P^\infty$. The rows of P^∞ are identical:

$$(P^\infty)_i{}^j = w^j.$$

Like the rows of P, the row vector w is a probability distribution: $w^j \geq 0$, $\sum_j w^j = 1$. w is the *equilibrium distribution* of the chain. The entry w^j tells the steady-state probability that the chain is in state j.

The row vector w is, up to multiplication by a constant, the unique row vector fixed by P:

$$\sum_i w^i P_i{}^j = w^j.$$

That makes it a row eigenvector corresponding to the eigenvalue 1. The corresponding column eigenvector is the constant vector:

$$\sum_j P_i{}^j \cdot 1 = 1.$$

Let M_{ij} be the expected time to get from state i to state j, where we take $M_{ii} = 0$. The *mean time from i to equilibrium* is

$$M_{iw} = \sum_j M_{ij} w^j.$$

This tells the expected time to get from state i to a state j selected randomly according to the equilibrium measure w.

John Kemeny (see [4, 4.4.10], [3, p. 469]) observed:

Theorem 1 M_{iw} *doesn't depend on i.*

The common value of the M_{iw}'s, denoted K, is the *Kemeny constant* or *seek time* of the chain.

Proof. Observe that the function M_{iw} is *discrete harmonic*, meaning that it has the *averaging property*

$$\sum_j P_i{}^j M_{jw} = M_{iw}.$$

The reason is that taking a step away from i brings you one step closer to your destination, except when your destination is i and the step begins a

2

wasted journey from i back to i: This happens with probability w^i, and the expected duration of the wasted journey is $\frac{1}{w^i}$, because the mean time between visits to i is the reciprocal of the equilibrium probability of being there. Thus

$$M_{iw} - 1 + w^i \frac{1}{w^i} = \sum_j P_i{}^j M_{jw},$$

so

$$M_{iw} = \sum_j P_i{}^j M_{jw}.$$

But now by the familiar *maximum principle*, any function f_i satisfying

$$\sum_j P_i{}^j f_j = f_i$$

must be constant: Choose i to maximize f_i, and observe that the maximum must be attained also for any j where $P_i{}^j > 0$; push the max around until it is attained everywhere. So M_{iw} doesn't depend on i. $\blacksquare$

Note. The application of the maximum principle we've made here shows that the only column eigenvectors having eigenvalue 1 for the matrix P are the constant vectors—a fact that was stated not quite explicitly above.

The foregoing argument shows the mean time from i to equilibrium is constant—but what is its value? For this we return to Kemeny's original proof of constancy for M_{iw}, which involved writing an explicit formula for M_{iw}, and noticing that it doesn't depend on i.

Define the *resolvent* or *fundamental matrix* or *Green's function*

$$
\begin{aligned}
Z &= (I - P^\infty) + (P - P^\infty) + (P^2 - P^\infty) + \ldots \\
&= (I - (P - P^\infty))^{-1} - P^\infty.
\end{aligned}
$$

Please be aware that this resolvent Z differs from the variant used by Kemeny and Snell [4] and Grinstead and Snell [3], which with our notation would be $(I - (P - P^\infty))^{-1}$. As others have observed (cf. Meyer [6]; Aldous and Fill [1]), for the version of Z we use here, the entries $Z_i{}^j$ have the natural probabilistic interpretation as the 'expected excess visits to j, starting from i, as compared with a chain started in equilibrium'. Accordingly we have

$$\sum_i w^i Z_i{}^j = 0$$

and

$$\sum_j Z_i{}^j = 0.$$

3

Since $Z_i{}^j$ measures excess visits to j starting at i, relative to starting in equilibrium, we obviously have $Z_j{}^j \geq Z_i{}^j$, because to make excess visits to j starting from i you first have to get to j. And the discrepancy $Z_j{}^j - Z_i{}^j$ is just $M_{ij}w^j$, because in equilibrium this is the expected number of visits to j over an interval of expected length M_{ij}. From this we get the familiar formula

$$M_{ij} = (Z_j{}^j - Z_i{}^j)\frac{1}{w^j}.$$

(Cf. [4, 4.4.7], [3, p. 459])

Proposition 2 (Kemeny and Snell [4, 4.4.10]) *Kemeny's constant is the trace of the resolvent Z:*

$$K = M_{iw} = \sum_j Z_j{}^j.$$

Proof.

$$
\begin{aligned}
M_{iw} &= \sum_j M_{ij}w^j \\
&= \sum_j Z_j{}^j - \sum_j Z_i{}^j \\
&= \sum_j Z_j{}^j,
\end{aligned}
$$

using the fact that $\sum_j Z_i{}^j = 0$. ■

This formula provides a computational verification that Kemeny's constant is constant, but doesn't explain *why* it is constant. Kemeny felt this keenly: A prize was offered for a more 'conceptual' proof, and awarded—rightly or wrongly—on the basis of the maximum principle argument outlined above.

Still, there are advantages to having an explicit formula. For starters, the explicit formula reveals that the seek time is a spectral invariant of the matrix $I - P$: If we denote the eigenvalues of $I - P$ by $\lambda_0 = 0, \lambda_1, \ldots, \lambda_{n-1}$, then the eigenvalues of $Z = (I - (P - P^\infty))^{-1} - P^\infty$ are $0, \frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_{n-1}}$, and

$$K = \text{Tr}(Z) = \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_{n-1}}.$$

(Cf. Meyer [6], Aldous and Fill [1], Levene and Loizou [5].) In terms of the eigenvalues $\alpha_0 = 1, \alpha_1 = 1 - \lambda_1, \ldots, \alpha_{n-1} = 1 - \lambda_{n-1}$ of P we have

$$K = \mathrm{Tr} Z = \frac{1}{1 - \alpha_1} + \ldots + \frac{1}{1 - \alpha_{n-1}}.$$

We'll have more to say about this later.

Kemeny on the bus

We now explore connections between the seek time, the famous *bus paradox*, and the Central Limit Theorem (CLT) for Markov Chains.

Just to recall, the bus paradox is that for a Poisson process like radioactive decay, the expected length of the interval between the events that bracket any given instant is twice the average interval between events—so if buses are dispatched by a Geiger counter, you must expect to wait twice as long for the bus as you would if the buses came at regular intervals. The explanation for this is that any given instant is more likely to land in a long inter-event interval than in a short inter-event interval, so we're taking a weighted average of the intervals between events, emphasizing the longer intervals, and this makes the expected waiting time longer than the average inter-event time. This inequality will be true for any renewal process in equilibrium: The factor of 2 disparity is special to the Poisson process, and arises because the Poisson process is memoryless and time-reversible.

To make the connection of the seek time to the bus paradox, we think about the *mean time from equilibrium to j*:

$$M_{wj} = \sum_i w^i M_{ij}.$$

Here we choose a starting state i at random according to w, and see how long it takes to get to j. This is backwards from what we did to define the seek time M_{iw}, where we looked at the time to get from i to equilibrium.

Unlike M_{iw}, which is independent of i, the quantity M_{wj} depends on the state j. Choosing the target state j at random according to w gets us back to K:

$$\sum_j M_{wj} w^j = K.$$

Proposition 3 (Kemeny and Snell [4, 4.4.9]) *The mean time from equilibrium to j is*

$$M_{wj} = Z_j{}^j \frac{1}{w^j}.$$

Proof.

$$M_{wj} = \sum_i w^i M_{ij} = \sum_i w^i (Z_j{}^j - Z_i{}^j) \frac{1}{w^j} = Z_j{}^j \frac{1}{w^j},$$

because $\sum_i w^i = 1$ and $\sum_i w^i Z_i{}^j = 0$. ■

To take advantage of this formula for M_{wj}, and specifically, to use it to derive the CLT for Markov chains, we now recall a bit of renewal theory. (Cf. Feller [2, Chapter XIII].)

A discrete renewal process is effectively just what you get if you watch a discrete-time Markov chain (possibly having infinitely many states) and take note of the times at which it is in some fixed state a. These times are called *renewal times* or *renewals*. The name derives from the fact that each time the chain reaches a it begins anew. The variables that tell the elapsed time between successive renewals are independent and identically distributed.

So let X be a random variable whose distribution is that for the time between successive renewals, and let it have mean μ and variance σ. We want to express the mean time τ from equilibrium to the next renewal in terms of μ and σ.

Proposition (The bus equality). For a discrete renewal process with interarrival times having mean μ and variance σ^2, the mean time τ from equilibrium to the next renewal is

$$\tau = \frac{\mu^2 + \sigma^2}{2\mu} - \frac{1}{2}.$$

The term $-\frac{1}{2}$ here is an artifact of using discrete time.

Proof. Let

$$p(n) = \mathrm{Prob}(X = n),$$

so

$$\mu = \mathrm{Exp}(X) = \sum_n n p(n)$$

and

$$\mu^2 + \sigma^2 = \mathrm{Exp}(X^2) = \sum_n n^2 p(n).$$

6

The expected time τ from equilibrium to the next renewal is

$$
\begin{aligned}
\tau &= \frac{\sum_n \frac{n-1}{2} n p(n)}{\sum_n n p(n)} \\
&= \frac{1}{2}\left(\frac{\sum_n n^2 p(n)}{\sum_n n p(n)} - 1\right) \\
&= \frac{\mu^2 + \sigma^2}{2\mu} - \frac{1}{2}. \quad \blacksquare
\end{aligned}
$$

Corollary 4 (The bus inequality)

$$
\tau \geq \frac{\mu}{2} - \frac{1}{2},
$$

with equality just if $\sigma = 0$ (all interarrival times equal). $\quad \blacksquare$

As in the bus equality above, the term $-\frac{1}{2}$ here is an artifact of using discrete time.

The bus equality shows that knowing the time τ from equilibrium to the next renewal is equivalent to knowing σ^2, the variance of the renewal time:

$$
\tau = \frac{\mu^2 + \sigma^2}{2\mu} - \frac{1}{2};
$$

$$
\sigma^2 = 2\mu\tau + \mu - \mu^2.
$$

Of course the mean renewal time μ is involved here, too: We take that for granted.

Now let's return to our Markov chain, and take for our renewal process visits to a given state j. The mean time between renewals is $\mu = \frac{1}{w^j}$. The expected time in equilibrium to the next renewal is $\tau = M_{wj} = Z_j{}^j \frac{1}{w^j}$. But above we saw that

$$
\sigma^2 = 2\mu\tau + \mu - \mu^2
$$

so

$$
\sigma^2 = 2Z_j{}^j \frac{1}{(w^j)^2} + \frac{1}{w^j} - \frac{1}{(w^j)^2}.
$$

Going back the other way, from μ and σ^2 we can find $Z_j{}^j$ (cf. Feller [2, (5.1) on p. 443]):

$$
Z_j{}^j = \frac{\sigma^2 - \mu + \mu^2}{2\mu^2}.
$$

Another piece of information about a renewal process that is equivalent to knowing σ^2 or τ (or $Z_j{}^j$, in the Markov chain case we just discussed) is the variance for the number of renewals over a long period, which shows up in the Central Limit Theorem for renewal processes:

Theorem 5 (CLT for renewal processes) *For a renewal process whose renewal time has mean μ and variance σ^2, the number of renewals over a long time T is approximately Gaussian with mean $T\frac{1}{\mu}$ and variance $T\frac{\sigma^2}{\mu^3}$.*

Idea of proof. To see a large number N of renewals will take time

$$T \approx N\mu \pm \sqrt{N}\sigma.$$

so the density of renewals over this interval is

$$
\begin{aligned}
\frac{N}{T} &= \frac{N}{N\mu \pm \sqrt{N}\sigma} \\
&= \frac{1}{\mu \pm \frac{1}{\sqrt{N}}\sigma} \\
&= \frac{1}{\mu(1 \pm \frac{1}{\sqrt{N}}\frac{\sigma}{\mu})} \\
&\approx \frac{1}{\mu}(1 \pm \frac{1}{\sqrt{N}}\frac{\sigma}{\mu}) \\
&= \frac{1}{\mu} \pm \frac{1}{\sqrt{N}}\frac{\sigma}{\mu^2} \\
&= \frac{1}{\mu} \pm \frac{1}{\sqrt{N\mu}}\frac{\sigma}{\mu^{\frac{3}{2}}} \\
&\approx \frac{1}{\mu} \pm \frac{1}{\sqrt{T}}\frac{\sigma}{\mu^{\frac{3}{2}}} \\
&= \frac{1}{\mu} \pm \frac{1}{\sqrt{T}}\sqrt{\frac{\sigma^2}{\mu^3}}.
\end{aligned}
$$

Thus

$$N \approx T\frac{1}{\mu} \pm \sqrt{T\frac{\sigma^2}{\mu^3}}. \quad \blacksquare$$

Note. Feller [2, p. 341] gives the following formulas for $\mathrm{Exp}(N)$ and $\mathrm{Exp}(N^2)$, which he derives—or rather, asks readers to derive—using generating functions:

$$\mathrm{Exp}(N) = \frac{T+1}{\mu} + \frac{\sigma^2 - \mu - \mu^2}{2\mu^2} + o(1).$$

$$\text{Exp}(N^2) = \frac{(T+2)(T+1)}{\mu^2} + \frac{2\sigma^2 - 2\mu - \mu^2}{\mu^3}T + o(T).$$

These combine to give

$$\text{Var}(N) = \text{Exp}(N^2) - Exp(N)^2 = \frac{\sigma^2}{\mu^3}T + o(T),$$

which is the same as what we get from the CLT.

The CLT for renewal processes translates into the following special case of the CLT for Markov chains (special because we are considering only the number of visits to one particular state, not the long-term average of a general function of the state).

Corollary 6 *For an ergodic Markov chain with resolvent Z, the number of visits to j over a long time is approximately Gaussian with mean $T\frac{1}{w^j}$ and variance*

$$T\frac{\sigma^2}{\mu^3} = T(2Z_j{}^j w^j + (w^j)^2 - w^j). \quad \blacksquare$$

Grinstead and Snell [3, p. 466] attribute this formula for the variance to Frechet.

Now just as in the case of the bus inequality, we get information from the fact that the variance here must be ≥ 0:

$$2Z_j{}^j w^j + (w^j)^2 - w^j \geq 0,$$

so

$$Z_j{}^j \geq \frac{1 - w^j}{2}.$$

Summing over j gives an inequality for the seek time K:

Proposition 7

$$K = \sum_j Z_j{}^j \geq \frac{n-1}{2}. \quad \blacksquare$$

This inequality for the seek time was observed by Levene and Loizou [5]. They derived it from the fact that if the non-1 eigenvalues of P are α_k, $1 \leq k \leq n-1$ then the non-1 eigenvalues of $I - P + P^\infty$ are $1 - \alpha_k$ and the non-0 eigenvalues of $Z = (I - P + P^\infty)^{-1} - P^\infty$ are $\lambda_k = \frac{1}{1-\alpha_k}$. But the α_k's lie in the unit disk, which maps to the region $\{x + iy : x \geq \frac{1}{2}\}$ under the map

taking z to $\frac{1}{1-z}$, so the non-0 eigenvalues of Z have real part $\geq \frac{1}{2}$, and thus K, which is real, satisfies

$$K = \mathrm{Tr}(Z) = \sum_{k=1}^{n-1} \lambda_k = \sum_{k=1}^{n-1} \frac{1}{1 - \alpha_k} \geq \frac{n-1}{2}.$$

Taking stock. From the resolvent Z we've computed the variance of the *return time*, meaning the time get from a designated starting state j back to j. If instead we're interested in the variance of a *hitting time*, meaning the time to get from i to $j \neq i$, we'll need to look at Z^2. We'd need this for the general CLT for Markov chains, which as noted above deals with the long-term average of a general function of the state of the chain, and requires knowing the covariance of the number of visits to a pair of states i and j. Looking beyond mean and variance, we get kth moments of return times from Z^{k-1} and kth moments of hitting times from Z^k. This was already in evidence for first moments: We need Z to find mean hitting times $M_{ij} = (Z_j{}^j - Z_i{}^j)\frac{1}{w^j}$, whereas expected return times $\frac{1}{w^j}$ don't require knowing Z at all.

References

[1] David Aldous and Jim Fill. Reversible markov chains and random walks on graphs. `http://www.stat.berkeley.edu/users/aldous/RWG/book.html`, 2002.

[2] William Feller. *An Introduction to Probability Theory and its Applications I, 3rd ed.* Wiley, 1968.

[3] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability: Second Revised Edition.* American Mathematical Society, 1997.

[4] John G. Kemeny and J. Laurie Snell. *Finite Markov chains.* Van Nostrand, 1960.

[5] Mark Levene and George Loizou. Kemeny's constant and the random surfer. *Amer. Math. Monthly*, 109(8):741–745, 2002.

[6] Carl D. Meyer, Jr. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Rev.*, 17:443–464, 1975.