Lecture 12

Markov Chains with Countably Many States

Classification of States

Definition (a) A state $j \in S = \mathbb{Z}_+$ (or $\mathbb{N}$, $\mathbb{Z}$ ...) is called *accessible* from state $i \in S$ if $p_{ij}^{(n)} > 0$ for some $n \geq 0$.
(b) The states $i \in S$ and $j \in S$ are said to *communicate* if i is accessible from j and j is accessible from i. We write $i \longleftrightarrow j$.

Remark Communication is an equivalence relation. The following holds.

- *Reflexivity*: $i \longleftrightarrow i, \quad i \in S$.
- *Symmetry*: $i \longleftrightarrow j \quad \Longleftrightarrow \quad j \longleftrightarrow i, \quad i, j \in S$.
- *Transitivity* $i \longleftrightarrow j$ and $j \longleftrightarrow k$ imply $i \longleftrightarrow k, \quad i, j, k \in S$.

As a consequence S decomposes into equivalence classes of communicating states.

Definition (a) A Markov chain $(X_n,\ n \geq 0)$ is called *irreducible* if S consists of precisely one such equivalence class, namely S itself.
(b) A state $i \in S$ is called *absorbing* if $p_{ii} = 1$.
(c) A state $i \in S$ is called *periodic* with period d if d is the smallest natural number such that for all n which are not a multiple of d we have $p_{ii}^{(n)} = 1$. If $d = 1$ then i is called *aperiodic*.

Remark All states in a given equivalence class have the same period.

Definition Transience and Recurrence (a) A state $i \in S$ is called *recurrent* if

$$f_i := P(X \text{ returns to } i | X_0 = i) = 1$$

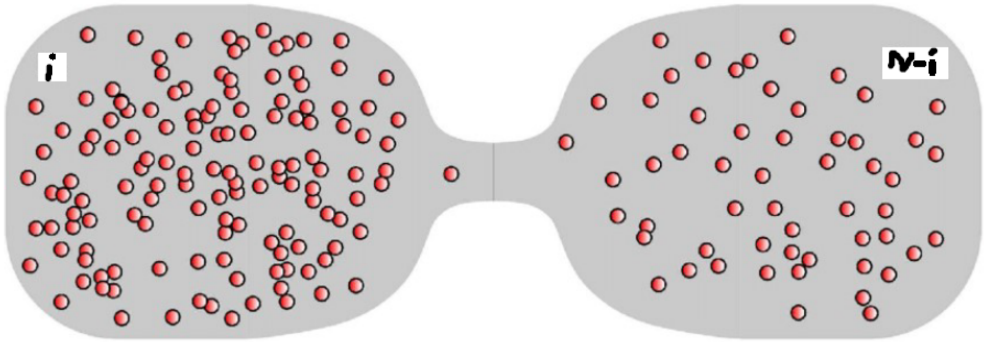
. (b) A state $i \in S$ is called *transient* if $f_i < 1$.

Proposition (a) A state $i \in S$ is recurrent if and only if $\sum_{n \in \mathbb{N}} p_{ii}^{(n)} = 1$, i. e.
(b) a state $i \in S$ is transient if and only if $\sum_{n \in \mathbb{N}} p_{ii}^{(n)} < 1$.

Definition Let $i \in S$ be recurrent and $T_i := \min\{n > 0 : X_n = i\}$.
(a) A state $i \in S$ is called *positive recurrent* if $E[T_i | X_0 = i] < \infty$.
(b) A state $i \in S$ is called *null recurrent* if $E[T_i | X_0 = i] = \infty$.

Proposition (a) In a given equivalence class all states are either transient or positive recurrent or null recurrent.
(b) In the case of just finitely many states, for example if $S = \{0, \ldots, M\}$, any recurrent state is positive recurrent.

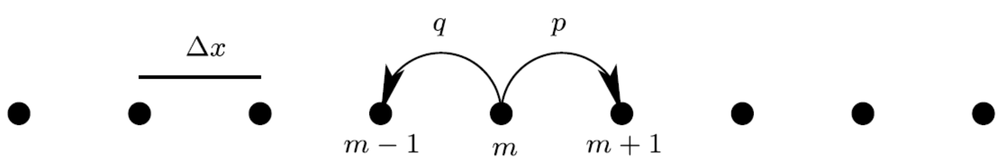
Example Ehrenfest model (see Lecture 2).



Marco Baldovin, research gate, modified

The Ehrenfest model has finitely many states $S = \{0, \ldots, M\}$, is irreducible and hence positive recurrent, and has period $d = 2$.

Example Symmetric random walk on $\mathbb{Z}$, that is $p_{i\,i-1} = p_{i\,i+1} = \frac{1}{2}$, $i \in \mathbb{Z}$, is null recurrent (left as advanced exercise) and has period $d = 2$.



course book, p. 3, $\Delta x = 1$, $p = q = \frac{1}{2}$

Stationarity and Limiting Distribution

As in the case with finitely many states we obtain for $S = \mathbb{Z}_+$ (or similarly for $S = \mathbb{N}$, $S = \mathbb{Z}$, ...)

$$\begin{aligned}\mathbf{p}^{(n+1)} &\equiv \left(p_0^{(n+1)}, p_1^{(n+1)}, \ldots\right) := \left(P\left(X_{n+1} = 0\right), P\left(X_{n+1} = 1\right), \ldots\right) \\ &= \mathbf{p}^{(n)} \cdot \mathbf{P} \quad \text{(multiplication of infinite line vector and infinite transition matrix)}\end{aligned}$$

which is meaningfully defined by

$$p_j^{(n+1)} = \sum_{i \in S} p_i^{(n)} p_{ij}\,, \quad j \in S.$$

Definition A distribution $\pi^* = (\pi_i^*,\ i \in S)$ is called *stationary* with respect to $(X_n,\ n \geq 0)$ if

$$\pi_j^* = \sum_{i \in S} \pi_i^* p_{ij}\,, \quad j \in S,$$

or short $\pi^* = \pi^* \cdot \mathbf{P}$.

Remarks (1) Existence and uniqueness of π^* is not clear without further analysis.

(2) If π^* exists then π^* is unique and it holds

$$E[T_i|X_0 = i] = \frac{1}{\pi_i^*}\,, \quad i \in S.$$

(3) If $\mathbf{p}^{(0)} = \pi^*$ then $\mathbf{p}^{(1)} = \pi^* \cdot \mathbf{P} = \pi^*$ and, in general,

$$\mathbf{p}^{(n)} = \pi^* \cdot \mathbf{P} \cdots \text{n times} \cdots \mathbf{P} = \pi^*\,, \quad n \in \mathbb{Z}_+.$$

Definition A distribution $\pi^\infty = (\pi_i^\infty,\ i \in S)$ is called *limiting distribution* of $(X_n,\ n \geq 0)$ if for arbitrary initial distribution $\mathbf{p}^{(0)} \equiv \left(p_0^{(0)}, p_1^{(0)}, \ldots\right) := (P(X_0 = 0), P(X_0 = 1), \ldots)$ it holds that

$$\lim_{n\to\infty} p_i^{(n)} = \pi_i^\infty\,, \quad i \in S.$$

Remarks (1) If π^∞ is a limiting distribution then π^∞ is also a stationary distribution since

$$\pi^\infty \mathbf{P} = \lim_{n\to\infty} \mathbf{p}^{(n)}\mathbf{P} = \lim_{n\to\infty} \mathbf{p}^{(n+1)} = \pi^\infty\,.$$

(2) If a limiting distribution π^∞ exists then, by definition, it is unique. In this case there exists a unique stationary distribution π^* and we have $\pi^\infty = \pi^*$. Sufficient conditions on the converse are formulated in the following theorem.

Theorem Let $X = (X_n,\ n \geq 0)$ be an irreducible and aperiodic Markov chain. Assume that there exists a stationary distribution π^*. Then π^* is the limiting distribution of $X = (X_n,\ n \geq 0)$, i. e. we have

$$\lim_{n\to\infty} p_i^{(n)} = \pi_i^*\,, \quad i \in S,$$

for any initial distribution $\mathbf{p}^{(0)}$.

Remark The crucial question is now the following. Under which condition does a stationary distribution π^* exist?

Theorem Let $X = (X_n,\ n \geq 0)$ be an irreducible and positive recurrent Markov chain. Then there exists a stationary distribution π^*.

Remark (Link to Lecture 2) An irreducible Markov chain with finitely many states is alway positive recurrent. If the chain is, in addition, aperiodic then it has a unique stationary distribution π^*.

Definition A Markov chain $X = (X_n,\ n \geq 0)$ is called *ergodic* if it is irreducible, aperiodic, and positive recurrent

Main Theorem, Ergodic Theorem An ergodic Markov chain $X = (X_n,\ n \geq 0)$ has a unique stationary distribution π^* which is, at the same time, the limiting distribution, that is,

$$\lim_{n\to\infty} p_i^{(n)} = \pi_i^*\,, \quad i \in S,$$

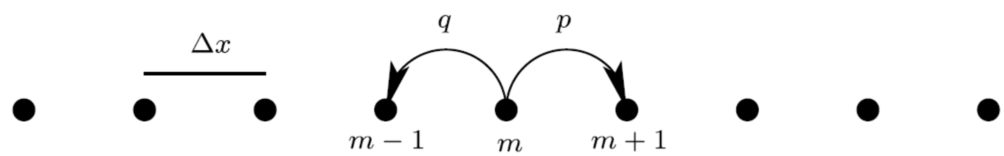
for any initial distribution $\mathbf{p}^{(0)}$. The probabilites π_i^*, $i \in S$, can be calculated as the unique solution of the system $\pi^* = \pi^*\mathbf{P}$ satisfying $\pi_i^* \geq 0$, $i \in S$, and $\sum_{i\in S} \pi_i^* = 1$.

From Random Walk to Diffusion on the Real Line

The following material has been taken from the course book, pp. 3-9. Even the way of writing has been taken over, for example $\mathbf{y}$ is a random variable with outcomes y, $\langle \mathbf{y} \rangle \equiv E[\mathbf{y}]$ and $p(m, N) \equiv p_{0\,m}^{(N)}$.

Step 1: The Markov Chain Model

Our aim is to answer the following question: What is the probability $p(m, N)$ that a not necessarily symmetric random walk will be at position m after N steps?



course book, p. 3

For $m < N$ there are many ways to start at 0, go through N jumps to nearest neighbor sites, and end up at m. But since all these possibilities are independent of each other we have to add up their probabilities. For all these ways we know that the random walk must have made $m + l$ jumps to the right and l jumps to the left; and since $m + 2l = N$, the random walk must have made

- $(N + m)/2$ jumps to the right and
- $(N - m)/2$ jumps to the left.

So whenever N is even, so is m. The probability for making exactly $(N + m)/2$ jumps to the right and exactly $(N - m)/2$ jumps to the left is

$$p^{\frac{1}{2}(N+m)} q^{\frac{1}{2}(N-m)} .$$

The number of ways to make $(N + m)/2$ out of N jumps to the right (and consequently $N - (N + m)/2 = (N - m)/2$ jumps to the left), where the order of the jumps does not matter (and repetitions are not allowed):

$$\frac{N!}{\left(\frac{N+m}{2}\right)! \left(\frac{N-m}{2}\right)!} .$$

The probability of being at position m after N jumps is therefore given as

$$p(m, N) = \frac{N!}{\left(\frac{N+m}{2}\right)! \left(\frac{N-m}{2}\right)!} p^{\frac{1}{2}(N+m)} (1 - p)^{\frac{1}{2}(N-m)},$$

which is the binomial distribution. If we know the probability distribution $p(m, N)$ we can calculate all the moments of m at fixed time N. Let us denote the number of jumps to the right as

$$r = (N + m)/2$$

and write

$$p(m, N) = p_N(r) = \frac{N!}{r!(N - r)!} p^r q^{N-r}$$

and calculate the moments of $p_N(r)$. For this purpose we use the property of the binomial distribution that $p_N(r)$ is the coefficient of u^r in $(pu+q)^N$. With this trick it is easy, for instance, to convince ourselves that $p_N(r)$ is properly normalized to one

$$\sum_{r=0}^{N} p_N(r) = \left[\sum_{r=0}^{N} \binom{N}{r} u^r p^r q^{N-r}\right]_{u=1} = \left[(pu+q)^N\right]_{u=1} = 1 \,.$$

The first moment or expectation value of the random variable $\mathbf{r}$ is:

$$\begin{aligned}
\langle \mathbf{r} \rangle &= \sum_{r=0}^{N} r p_N(r) \\
&= \left[\sum_{r=0}^{N} r \binom{N}{r} u^r p^r q^{N-r}\right]_{u=1} = \left[\sum_{r=0}^{N} \binom{N}{r} u \frac{\mathrm{d}}{\mathrm{d}u}\left(u^r p^r q^{N-r}\right)\right]_{u=1} \\
&= \left[u \frac{\mathrm{d}}{\mathrm{d}u} \sum_{r=0}^{N} \binom{N}{r} u^r p^r q^{N-r}\right]_{u=1} = \left[u \frac{\mathrm{d}}{\mathrm{d}u}(pu+q)^N\right]_{u=1} \\
&= \left[Nup(pu+q)^{N-1}\right]_{u=1} \,.
\end{aligned}$$

leading to

$$\mathrm{E}[\mathbf{r}] \equiv \langle \mathbf{r} \rangle = Np \,.$$

In the same manner, one can derive the following for the second moment.

$$\mathrm{E}\left[\mathbf{r}^2\right] \equiv \left\langle \mathbf{r}^2 \right\rangle = \left[\left(u \frac{\mathrm{d}}{\mathrm{d}u}\right)^2 (pu+q)^N\right]_{u=1} = Np + N(N-1)p^2 \,.$$

From this one can calculate the variance or second central moment

$$\mathrm{Var}[\mathbf{r}] \equiv \sigma_{\mathbf{r}}^2 := \left\langle (\mathbf{r} - \langle \mathbf{r} \rangle)^2 \right\rangle = \left\langle \mathbf{r}^2 \right\rangle - \langle \mathbf{r} \rangle^2$$

of the distribution, which is a measure of the width of the distribution

$$\sigma_{\mathbf{r}}^2 = Npq \,.$$

The relative width of the distribution

$$\frac{\sigma_{\mathbf{r}}}{\langle \mathbf{r} \rangle} = \sqrt{\frac{q}{p}} N^{-1/2}$$

goes to zero with increasing number of performed steps, N. Recalling

$$m = 2r - N$$

we get the following results

$$\langle \mathbf{m} \rangle = 2N\left(p - \frac{1}{2}\right)$$

and

$$\left\langle \mathbf{m}^2 \right\rangle = 4Np(1-p) + 4N^2\left(p - \frac{1}{2}\right)^2$$
$$\sigma^2 = \left\langle \mathbf{m}^2 \right\rangle - \langle \mathbf{m} \rangle^2 = 4Npq \,.$$

In the case of symmetric jump rates, this reduces to

$$\langle \mathbf{m} \rangle = 0 \quad \text{and} \quad \langle \mathbf{m}^2 \rangle = N\,.$$

Step 2: The Markov chain Model for large N

Assuming $N \gg 1$ we can use Stirling's formula to approximate the factorials in the binomial distribution

$$\ln N! = \left(N + \frac{1}{2}\right) \ln N - N + \frac{1}{2} \ln 2\pi + \mathrm{O}\left(N^{-1}\right)$$

Using Stirling's formula, we get

$$\begin{aligned}
\ln p(m, N) = &\left(N + \frac{1}{2}\right) \ln N - \left(\frac{N+m}{2} + \frac{1}{2}\right) \ln \left[\frac{N}{2}\left(1 + \frac{m}{N}\right)\right] \\
&- \left(\frac{N-m}{2} + \frac{1}{2}\right) \ln \left[\frac{N}{2}\left(1 - \frac{m}{N}\right)\right] \\
&+ \frac{N+m}{2} \ln p + \frac{N-m}{2} \ln q - \frac{1}{2} \ln 2\pi\,.
\end{aligned}$$

Now we want to derive an approximation to the binomial distribution close to its maximum, which is also close to the expectation value $\langle \mathbf{m} \rangle$. So let us write

$$m = \langle \mathbf{m} \rangle + \delta m = 2Np - N + \delta m$$

which leads to

$$\frac{N+m}{2} = Np + \frac{\delta m}{2} \quad \text{and} \quad \frac{N-m}{2} = Nq - \frac{\delta m}{2}\,.$$

Using these relations, we get

$$\begin{aligned}
\ln p(m, N) = &\left(N + \frac{1}{2}\right) \ln N - \frac{1}{2} \ln 2\pi \\
&+ \left(Np + \frac{\delta m}{2}\right) \ln p + \left(Nq - \frac{\delta m}{2}\right) \ln q \\
&- \left(Np + \frac{\delta m}{2} + \frac{1}{2}\right) \ln \left[Np\left(1 + \frac{\delta m}{2Np}\right)\right] \\
&- \left(Nq - \frac{\delta m}{2} + \frac{1}{2}\right) \ln \left[Nq\left(1 - \frac{\delta m}{2Nq}\right)\right] \\
= &-\frac{1}{2} \ln(2\pi Npq) - \left(Np + \frac{\delta m}{2} + \frac{1}{2}\right) \ln \left(1 + \frac{\delta m}{2Np}\right) \\
&- \left(Nq - \frac{\delta m}{2} + \frac{1}{2}\right) \ln \left(1 - \frac{\delta m}{2Nq}\right)\,.
\end{aligned}$$

Expanding the logarithm

$$\ln(1 \pm x) = \pm x - \frac{1}{2} x^2 + \mathrm{O}\left(x^3\right)$$

yields

$$\ln p(m, N) \simeq -\frac{1}{2}\ln(2\pi Npq) - \frac{1}{2}\frac{(\delta m)^2}{4Npq} - \frac{\delta m(q-p)}{4Npq} + \frac{(\delta m)^2\,(p^2+q^2)}{16(Npq)^2}\,.$$

We recall that the variance (squared width) of the binomial distribution is $\sigma^2 = 4Npq$. When we want to approximate the distribution in its center and up to fluctuations around the mean value of the order $(\delta m)^2 = \mathrm{O}\left(\sigma^2\right)$, we find for the last terms in the above equation:

$$\frac{\delta m(q-p)}{4Npq} = \mathrm{O}\left((Np)^{-1/2}\right) \text{ and } \frac{(\delta m)^2\,(p^2+q^2)}{16(Npq)^2} = \mathrm{O}\left((Np)^{-1}\right)\,.$$

These terms can be neglected if $Np \to \infty$. We therefore obtain

$$p(m, N) \to \frac{2}{\sqrt{2\pi 4Npq}}\exp\left[-\frac{1}{2}\frac{(\delta m)^2}{4Npq}\right]\,.$$

Step 3: Deriving the Density of the Diffusion and the Diffusion PDE by Scaling

Let us write

$$\begin{aligned} x &= m\Delta x, \qquad \text{i.e., } \langle\mathbf{x}\rangle = \langle\mathbf{m}\rangle\Delta x \\ t &= N\Delta t \\ D &= 2pq\frac{(\Delta x)^2}{\Delta t} \end{aligned}$$

so that we can interpret

$$p(m\Delta x, N\Delta t) = \frac{2\Delta x}{\sqrt{2\pi 2Dt}}\exp\left[-\frac{1}{2}\frac{(x-\langle\mathbf{x}\rangle)^2}{2Dt}\right]$$

as the probability of finding the random walk in an interval of width $2\Delta x$ around a certain position x at time t. We now require that

$$\Delta x \to 0, \quad \Delta t \to 0, \quad \text{and} \quad 2pq\frac{(\Delta x)^2}{\Delta t} = D = \text{const.}$$

Here, D, with the units length $^2/$ time, is called the *diffusion coefficient* of the walk. For the probability that the random walk is in an interval of width $\mathrm{d}x$ around the position x we get

$$p(x, t)\mathrm{d}x = \frac{1}{\sqrt{2\pi 2Dt}}\exp\left[-\frac{1}{2}\frac{(x-\langle\mathbf{x}\rangle)^2}{2Dt}\right]\mathrm{d}x\,.$$

When we look closer at the definition of $\langle\mathbf{x}\rangle$ above, we see that we have used another assumption in the limiting procedure:

$$\langle\mathbf{x}\rangle(t) = \Delta x\langle\mathbf{m}\rangle = 2\left(p-\frac{1}{2}\right)N\Delta x = 2\left(p-\frac{1}{2}\right)\frac{\Delta x}{\Delta t}t\,.$$

So our limiting procedure also has to include the requirement

$$\Delta x \to 0, \quad \Delta t \to 0 \quad \text{and} \quad \frac{2\left(p-\frac{1}{2}\right)\Delta x}{\Delta t} = v = \text{ const.}$$

As already discussed before, when $p = 1/2$ the average position of the walk is at zero for all times and the velocity of the walk vanishes. Any asymmetry in the transition rates ($p \neq q$) produces a velocity of the walk. However, when $v = 0$ we have $\langle \mathbf{x} \rangle = 0$ and $\langle \mathbf{x}^2 \rangle = 2Dt$. We can write down the probability density for the position of the random walk at time t,

$$p(x,t) = \frac{1}{\sqrt{2\pi 2Dt}} \exp\left[-\frac{1}{2}\frac{(x-vt)^2}{2Dt}\right]$$

with starting condition

$$p(x,0) = \delta(x)$$

and boundary conditions

$$p(x,t) \stackrel{x\to\pm\infty}{\longrightarrow} 0\,.$$

By substitution one can confirm that the above density is the solution of the following partial differential equation:

$$\frac{\partial}{\partial t}p(x,t) = -v\frac{\partial}{\partial x}p(x,t) + D\frac{\partial^2}{\partial x^2}p(x,t)$$

which is *Fick's equation for diffusion in the presence of a constant drift.*

Rayleigh-Pearson Walk

The material has been taken from the course book, pp. 69-71. It concerns Lord Rayleigh's contribution on the two-dimensional random walk: 'What is the probability for the random walk to be at a distance between r and $r + \mathrm{d}r$ from his starting position after n steps?' The solution is

$$p(r)\mathrm{d}r = \frac{2}{n}\mathrm{e}^{-r^2/n} r\ \mathrm{d}r.$$

To characterize the random walk we note that

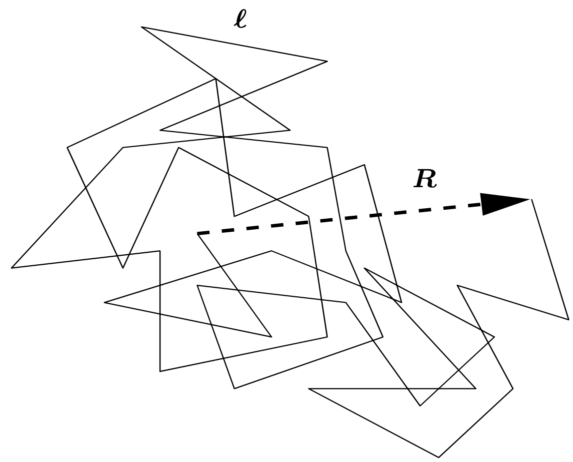
- the steps occur at regularly spaced time points (discrete time),
- the walk is isotropic, and
- the jump distance has a distribution $p(\ell)$?

For dimension d one even gets for Fourier transform of the density

$$\begin{aligned}\widehat{p_n}(r) &= \frac{1}{(2\pi)^d}\int \mathrm{d}^d k \mathrm{e}^{-(n/2d)k^2\langle \ell^2\rangle}\mathrm{e}^{-\mathrm{i}k\cdot r} \\ &= \left(\frac{d}{2\pi n\langle \ell^2\rangle}\right)^{d/2}\exp\left[-\frac{dr^2}{2n\langle \ell^2\rangle}\right]\,.\end{aligned}$$

Specializing to dimension 2 as well as $\langle \ell^2 \rangle = 1$, we find Lord Rayleigh's result,

$$p_n(r) = 2\pi r\widehat{p_n}(r) = \frac{2r}{n}\mathrm{e}^{-r^2/n}\,.$$



from course book, p. 71

A Polymer Model

The material has been taken from the course book, pp. 71-72. The Rayleigh-Pearson random walk also appears in a model in polymer physics. It describes a polymer as a series of links of fixed length $\ell = \ell_0$, which are connected through random angles, and crossings of these links in space are allowed. This is the Rayleigh-Pearson walk in $d = 3$ with $p(\ell) = \delta(\ell - \ell_0)$ and the number of the repeat unit or monomer when we go along the chain replacing time. For the length distribution of the end-to-end vector of a chain of N links, we can write

$$\begin{aligned} p_N(R) &= 4\pi R^2 \widehat{p_N}(R) \\ &= \sqrt{2/\pi} \left(\frac{N\ell_0^2}{3} \right)^{-3/2} R^2 \exp\left[-\frac{3R^2}{2N\ell_0^2} \right] \end{aligned}$$

From this, we find that the mean-square end-to-end distance of the polymer chain is

$$\begin{aligned} \langle R^2 \rangle &= \int_0^\infty \mathrm{d}R \sqrt{2/\pi} \left(\frac{N\ell_0^2}{3} \right)^{-3/2} R^4 \exp\left[-\frac{3R^2}{2N\ell_0^2} \right] \\ &= N\ell_0^2 . \end{aligned}$$