POLAND-SCHERAGA model and renewal theory

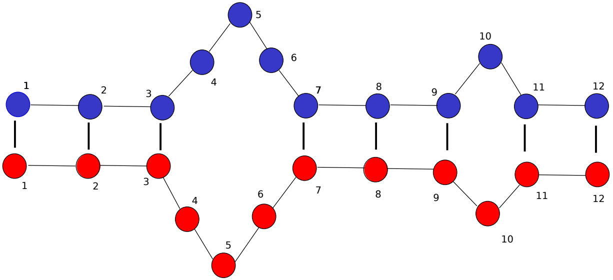
Maha KHATIB

Supervised by Giambattista GIACOMIN
LPMA - University of Paris 7-Denis Diderot

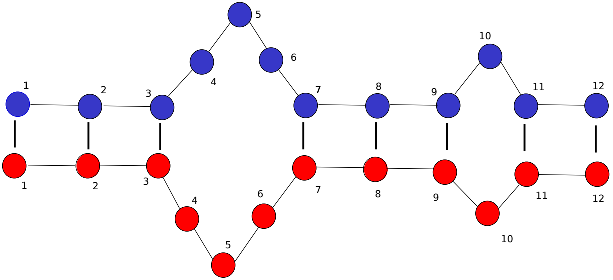
22 Avril 2016
Colloque Jeunes Probabilistes et Statisticiens

Plan

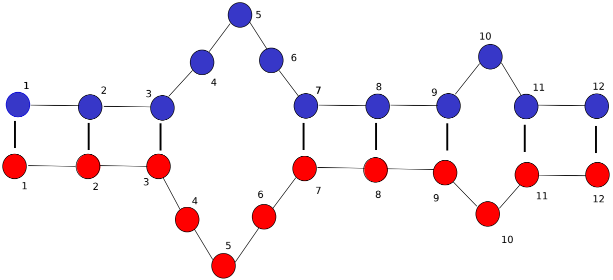
- A DNA molecule is composed of an alternating sequence of bound and denaturated states.



- A DNA molecule is composed of an alternating sequence of bound and denaturated states.

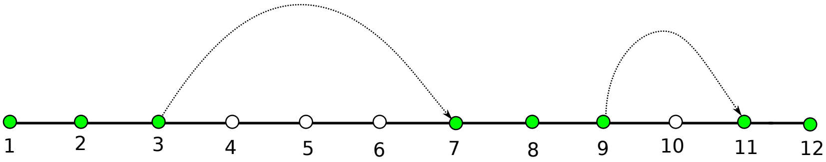


- A DNA molecule is composed of an alternating sequence of bound and denaturated states.



- The statistical weight:
  - bound sequence of length k: $\exp(-kE_b/T)$.
  - loop of length k: As^k/k^c.

- Link with the pinning model ?



- A discrete renewal issued from the origin is a random walk $\tau = \{\tau_n\}_{n=0,1,\dots}$ where $\tau_0 = 0$ and, for $n \in \mathbb{N}$, τ_n is a sum of n independent identically distributed random variables taking values in $\mathbb{N}^2$.

- A discrete renewal issued from the origin is a random walk $\tau = \{\tau_n\}_{n=0,1,\ldots}$ where $\tau_0 = 0$ and, for $n \in \mathbb{N}$, τ_n is a sum of n independent identically distributed random variables taking values in $\mathbb{N}^2$.
- Let

$$\mathbf{P}(\tau_1 = n) = K(n) := \frac{L(n)}{n^{1+\alpha}},$$

where $L(\cdot)$ is a slowly varying function and $\alpha > 0$.

- A discrete renewal issued from the origin is a random walk $\tau = \{\tau_n\}_{n=0,1,\ldots}$ where $\tau_0 = 0$ and, for $n \in \mathbb{N}$, τ_n is a sum of n independent identically distributed random variables taking values in $\mathbb{N}^2$.
- Let

$$\mathbf{P}(\tau_1 = n) = K(n) := \frac{L(n)}{n^{1+\alpha}},$$

where $L(\cdot)$ is a slowly varying function and $\alpha > 0$.
- Without loss of generality, we suppose that

$$\sum_{n \geq 1} K(n) = 1,$$

- The polymer measure $\mathbf{P}^c_{N,h}$ is defined as

$$\frac{\mathrm{d}\mathbf{P}^c_{N,h}}{\mathrm{d}\mathbf{P}} := \frac{1}{Z^c_{N,h}} \exp\left(h \sum_{n=1}^{N} \mathbf{1}_{n\in\tau} \right) \mathbf{1}_{N\in\tau} ,$$

- The polymer measure $\mathbf{P}^c_{N,h}$ is defined as

$$\frac{\mathrm{d}\mathbf{P}^c_{N,h}}{\mathrm{d}\mathbf{P}} := \frac{1}{Z^c_{N,h}} \exp\left(h \sum_{n=1}^{N} \mathbf{1}_{n\in\tau} \right) \mathbf{1}_{N\in\tau} \,,$$

- The partition function

$$Z^c_{N,h} = \sum_{n=1}^{N} \sum_{\substack{\underline{l}\in\mathbb{N}^n: \\ |\underline{l}|=N}} \prod_{i=1}^{n} \exp(h) K\left(l_i\right) = \exp\left(NF(h)\right) \mathbf{P}\left(N \in \widetilde{\tau}_h\right) \,,$$

- The polymer measure $\mathbf{P}^c_{N,h}$ is defined as

$$\frac{\mathrm{d}\mathbf{P}^c_{N,h}}{\mathrm{d}\mathbf{P}} := \frac{1}{Z^c_{N,h}} \exp\left(h \sum_{n=1}^{N} \mathbf{1}_{n\in\tau} \right) \mathbf{1}_{N\in\tau} \,,$$

- The partition function

$$Z^c_{N,h} = \sum_{n=1}^{N} \sum_{\substack{\underline{l}\in\mathbb{N}^n: \\ |\underline{l}|=N}} \prod_{i=1}^{n} \exp(h) K\left(l_i\right) = \exp\left(NF(h)\right) \mathbf{P}\left(N \in \widetilde{\tau}_h\right) \,,$$

- with

$$\mathbf{P}(\widetilde{\tau}_h = n) = K(n) \exp(-F(h)n + h) \,,$$

- The polymer measure $\mathbf{P}^c_{N,h}$ is defined as

$$\frac{\mathrm{d}\mathbf{P}^c_{N,h}}{\mathrm{d}\mathbf{P}} := \frac{1}{Z^c_{N,h}} \exp\left(h \sum_{n=1}^{N} \mathbf{1}_{n \in \tau} \right) \mathbf{1}_{N \in \tau} ,$$

- The partition function

$$Z^c_{N,h} = \sum_{n=1}^{N} \sum_{\substack{\underline{l} \in \mathbb{N}^n : \\ |\underline{l}| = N}} \prod_{i=1}^{n} \exp(h) K(l_i) = \exp\left(NF(h)\right) \mathbf{P}\left(N \in \widetilde{\tau}_h\right) ,$$

- with

$$\mathbf{P}(\widetilde{\tau}_h = n) = K(n) \exp(-F(h)n + h) ,$$

- $F(\cdot)$ is the free energy: the unique solution of

$$\sum_{n \geq 1} K(n) \exp\left(-F(h)n + h\right) = 1 .$$

- Renewal theorem:

$$\lim_{N\to\infty} \mathbf{P}\left(N \in \widetilde{\tau}_h\right) = 1/\mathbf{E}[\widetilde{\tau}_{h,1}] .$$

- Renewal theorem:

$$\lim_{N\to\infty} \mathbf{P}\left(N \in \widetilde{\tau}_h\right) = 1/\mathbf{E}[\widetilde{\tau}_{h,1}]\,.$$

- If $h = 0$ and $\mathbf{E}[\widetilde{\tau}_{0,1}] = \sum_{n\geq 1} nK(n) = \infty$ (implied by $\alpha \in (0,1)$) then

$$\mathbf{P}\left(N \in \widetilde{\tau}_h\right) \overset{N\to\infty}{\sim} \frac{\alpha \sin(\Pi\alpha)}{\Pi} \frac{1}{N^{1-\alpha}}\,.$$

- Renewal theorem:

$$\lim_{N\to\infty} \mathbf{P}\left(N \in \widetilde{\tau}_h\right) = 1/\mathbf{E}[\widetilde{\tau}_{h,1}]\,.$$

- If $h = 0$ and $\mathbf{E}[\widetilde{\tau}_{0,1}] = \sum_{n\geq 1} nK(n) = \infty$ (implied by $\alpha \in (0,1)$) then

$$\mathbf{P}\left(N \in \widetilde{\tau}_h\right) \overset{N\to\infty}{\sim} \frac{\alpha \sin(\Pi\alpha)}{\Pi} \frac{1}{N^{1-\alpha}}\,.$$

- The free energy

$$F(h) = \lim_{N\to\infty} \frac{1}{N} \log Z^{c}_{N,h}\,.$$

- We define the critical point

$$h_c := \sup\{h \, : \, F(h) = 0\}\,, \tag{1}$$

at which a localization/ delocalization transition takes place in the system.

- We define the critical point

$$h_c := \sup\{h \,:\, F(h) = 0\}\,, \tag{1}$$

at which a localization/ delocalization transition takes place in the system.

Theorem

For every choice of $\alpha \geq 0$ and $L(\cdot)$, there exists a slowly varying function $\hat{L}(\cdot)$ such that

$$F(h) = h^{1/\min(1,\alpha)} \hat{L}(1/h)\,.$$

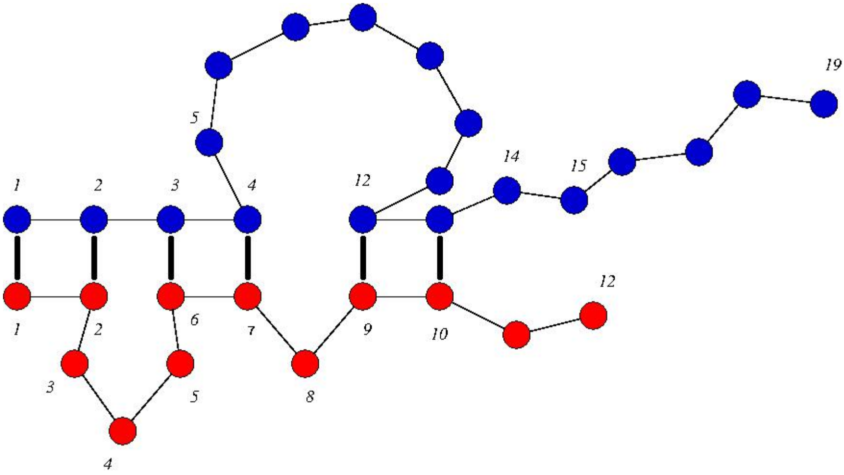
- The generalized Poland-Scheraga model:
  The possibility of formation of non-symmetrical loops in the two strands (i.e., the contribution to a loop, in terms of number of bases, from the two strands is not necessarily the same).

- The generalized Poland-Scheraga model:
  The possibility of formation of non-symmetrical loops in the two strands (i.e., the contribution to a loop, in terms of number of bases, from the two strands is not necessarily the same).
- The two strands may be of different lengths.

- The generalized Poland-Scheraga model:
  The possibility of formation of non-symmetrical loops in the two strands (i.e., the contribution to a loop, in terms of number of bases, from the two strands is not necessarily the same).
- The two strands may be of different lengths.
- 1. Each base pair is energetically favored and carries an energy $-E_b < 0$;
  2. A base which is not in pair is in a loop: we associate to a loop of length ℓ an entropy factor

$$B(\ell) := s^{\ell} \ell^{-c} \,, \tag{2}$$

where $s \geq 1$ and $c > 2$.


1
2
3
4
5
12
14
15
19
1
2
3
4
5
6
7
8
9
10
12

- Bivariate renewal theory:
  A discrete two-dimensional renewal issued from the origin is a random walk
  $\tau = \{\tau_n\}_{n=0,1,\ldots} = (\tau^{(1)}, \tau^{(2)}) = \{(\tau_n^{(1)}, \tau_n^{(2)})\}_{n=0,1,\ldots}$ where $\tau_0 = (0,0)$ and, for $n \in \mathbb{N} := \{1, 2, \ldots\}$.

- Bivariate renewal theory:
  A discrete two-dimensional renewal issued from the origin is a random walk
  $\tau = \{\tau_n\}_{n=0,1,\ldots} = (\tau^{(1)}, \tau^{(2)}) = \{(\tau_n^{(1)}, \tau_n^{(2)})\}_{n=0,1,\ldots}$ where $\tau_0 = (0,0)$ and, for $n \in \mathbb{N} := \{1, 2, \ldots\}$.
- We set

$$\mathrm{K}(n,m) := \mathbf{P}(\tau_1 = (n,m)) = \frac{L(n+m)}{(n+m)^{1+\alpha}},$$

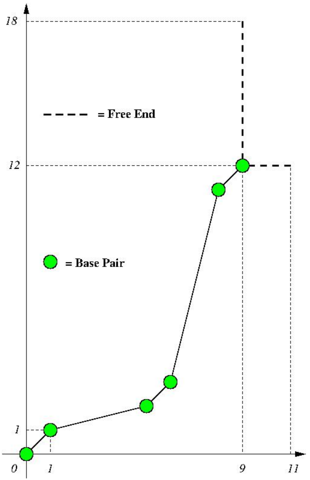
with $\alpha \geq 1$.

- Bivariate renewal theory:
  A discrete two-dimensional renewal issued from the origin is a random walk
  $\tau = \{\tau_n\}_{n=0,1,\ldots} = (\tau^{(1)}, \tau^{(2)}) = \{(\tau_n^{(1)}, \tau_n^{(2)})\}_{n=0,1,\ldots}$ where $\tau_0 = (0,0)$ and, for $n \in \mathbb{N} := \{1,2,\ldots\}$.
- We set

$$\mathrm{K}(n,m) := \mathbf{P}(\tau_1 = (n,m)) = \frac{L(n+m)}{(n+m)^{1+\alpha}},$$

  with $\alpha \geq 1$.
- The polymer measure $\mathbf{P}^c_{N,M,h}$ is defined as

$$\frac{\mathrm{d}\mathbf{P}^c_{N,M,h}}{\mathrm{d}\mathbf{P}} := \frac{1}{Z^c_{N,M,h}} \exp\left(h \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbf{1}_{(n,m)\in\tau} \right) \mathbf{1}_{(N,M)\in\tau}.$$


18
= Free End
12
= Base Pair
1
0
1
9
11

- The partition function

$$Z^c_{N,M,h} = \exp\left((N+M)\mathrm{G}\right) \sum_{n=1}^{N\wedge M} \sum_{\substack{\underline{l}\in\mathbb{N}^n:\\ |\underline{l}|=N}} \sum_{\substack{\underline{t}\in\mathbb{N}^n:\\ |\underline{t}|=M}} \prod_{i=1}^{n} \widetilde{K}_h\left(l_i, t_i\right)$$

$$= \exp\left((N+M)\mathrm{G}\right) \mathbf{P}\left((N,M) \in \widetilde{\tau}_h\right),$$

- The partition function

$$Z^c_{N,M,h} = \exp\left((N+M)\mathrm{G}\right) \sum_{n=1}^{N\wedge M} \sum_{\substack{\underline{l}\in\mathbb{N}^n:\\ |\underline{l}|=N}} \sum_{\substack{\underline{t}\in\mathbb{N}^n:\\ |\underline{t}|=M}} \prod_{i=1}^{n} \widetilde{K}_h\left(l_i, t_i\right)$$

$$= \exp\left((N+M)\mathrm{G}\right) \mathbf{P}\left((N,M)\in\widetilde{\tau}_h\right) ,$$

- where $\widetilde{K}_h(n,m) = \exp(h-(n+m)\mathrm{G})K(n+m)$ and $\mathrm{G} = \mathrm{G}(h)$ is the only solution to

$$\sum_{n,m} K(n+m)\exp(h-(n+m)\mathrm{G}) = 1 , \qquad (3)$$

when such a solution exists (that is, when $h \geq 0$), and $\mathrm{G} = 0$ otherwise.

Theorem (Borovkov, Mogulskil 1996)

For every $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$

$$\lim_{t \to \infty} \frac{1}{t} \log \mathbf{P}\left(\lceil t\theta \rceil \in \tau\right) = -D\left(\theta\right),$$

where

$$D(\theta) = \sup_{\lambda \in A} \langle \lambda, \theta \rangle = \sup_{\lambda \in \partial A} \langle \lambda, \theta \rangle,$$

and A *is the closed convex set* $\{\lambda \in \mathbb{R}^2 : \mathbf{E}[\exp(\langle \lambda, \tau \rangle)] \leq 1\}$ *and* ∂A *is the boundary of* A.

Theorem

For every $\gamma \geq 1$ *,*

$$\mathrm{F}_\gamma(h) := \lim_{\substack{N,M\to\infty:\\ \frac{M}{N}\to\gamma}} \frac{1}{N} \log Z^c_{N,M,h}$$

with

$$\mathrm{F}_\gamma(h) = \begin{cases} 0 & \text{if } h \leq 0\,, \\ (1+\gamma)\mathrm{G}(h) - D_h(1,\gamma) & \text{if } h > 0\,, \end{cases}$$

- The critical point is

$$h_c := \inf\{h : \mathrm{F}_\gamma(h) > 0\} = \max\{h : \mathrm{F}_\gamma(h) = 0\} = 0\,,$$

- We have

$$D_h(1,\gamma) = \max_{\lambda \in B_h} \left(\lambda_1 + \gamma\lambda_2\right) ,$$

where

$$B_h = \left\{ \lambda : \sum_{n,m} K(n+m) e^{h - \mathrm{G}(h)(n+m)} e^{\lambda_1 n + \lambda_2 m} = 1 \right\} .$$

- We have

$$D_h(1,\gamma) = \max_{\lambda \in B_h} (\lambda_1 + \gamma\lambda_2) ,$$

where

$$B_h = \left\{ \lambda : \sum_{n,m} K(n+m) e^{h - \mathrm{G}(h)(n+m)} e^{\lambda_1 n + \lambda_2 m} = 1 \right\} .$$

- Let

$$\overline{\lambda}_1(h) := \sup \left\{ \lambda_1 < 0 : \sum_{n,m} K(n+m) e^{h - (\mathrm{G}(h) - \lambda_1) n} \leq 1 \right\} ,$$

then

$$\overline{\lambda}_1 \leq \lambda_1 \leq 0,\ 0 \leq \lambda_2 \leq \mathrm{G}. \qquad (4)$$

- We have

$$D_h(1,\gamma) = \max_{\lambda \in B_h} (\lambda_1 + \gamma\lambda_2) ,$$

where

$$B_h = \left\{ \lambda : \sum_{n,m} K(n+m) e^{h - \mathrm{G}(h)(n+m)} e^{\lambda_1 n + \lambda_2 m} = 1 \right\} .$$

- Let

$$\overline{\lambda}_1(h) := \sup \left\{ \lambda_1 < 0 : \sum_{n,m} K(n+m) e^{h - (\mathrm{G}(h) - \lambda_1) n} \le 1 \right\} ,$$

then

$$\overline{\lambda}_1 \le \lambda_1 \le 0,\ 0 \le \lambda_2 \le \mathrm{G}. \tag{4}$$

- For $\gamma = 1$, we have $D_h(1,1) = 0$. Then

$$\mathrm{F}_1(h) = 2\mathrm{G}(h) .$$

Theorem

For every choice of $\alpha > 1$ and $L(\cdot)$, there exists a slowly varying function $\hat{L}(\cdot)$ such that

$$\mathrm{F}_1(h) = h^{1/\min(1,\alpha-1)}\hat{L}(1/h)\,.$$

Moreover there exists $c_{\alpha,\gamma}$ such that

$$\mathrm{F}_\gamma(h) \overset{h \searrow 0}{\sim} c_{\alpha,\gamma}\mathrm{F}_1(h)\,.$$

- We set, always for $h > 0$

$$\gamma_c(h) \ := \frac{\sum_{n,m} m\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)}{\sum_{n,m} n\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)} \,.$$

- We set, always for $h > 0$

$$\gamma_c(h) := \frac{\sum_{n,m} m\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)}{\sum_{n,m} n\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)} .$$

- In the Cramér regime (the maximum is achieved in the interior: when $\gamma_c(h) > \gamma$)

$$\mathrm{F}_\gamma(h) = \left(\mathrm{G}(h) - \hat{\lambda}_1(h)\right) + \gamma\left(\mathrm{G}(h) - \hat{\lambda}_2(h)\right) .$$

- We set, always for $h > 0$

$$\gamma_c(h) := \frac{\sum_{n,m} m\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)}{\sum_{n,m} n\, K(n+m) \exp\left(-n\left(\mathrm{G}(h) - \overline{\lambda}_1(h)\right)\right)}\,.$$

- In the Cramér regime (the maximum is achieved in the interior: when $\gamma_c(h) > \gamma$)

$$\mathrm{F}_\gamma(h) = \left(\mathrm{G}(h) - \hat{\lambda}_1(h)\right) + \gamma\left(\mathrm{G}(h) - \hat{\lambda}_2(h)\right)\,.$$

- Out of the Cramér regime (the maximum is achieved at the boundary: when $\gamma_c(h) \leq \gamma$)

$$\mathrm{F}_\gamma(h) = \mathrm{G}(h) - \overline{\lambda}_1(h)\,.$$

Theorem

Fix $\gamma \geq 1$*.* $\mathrm{F}_\gamma(\cdot)$ *is analytic on* $\{h:\ h>0$ *such that* $\gamma_c(h)-\gamma \neq 0\}$ *and* $\mathrm{F}_\gamma(\cdot)$ *is not analytic for the values* $h>0$ *at which* $\gamma_c(h)-\gamma$ *changes sign. However,* $\mathrm{F}'_\gamma(\cdot)$ *is continuous on the positive semi-axis.*

Thank you for your attention :)