Targeted cutting of random recursive trees.

Sergio I. López

joint work with Laura Eslava, Marco L. Ortiz.
(Universidad Nacional Autónoma de México),

Probability Seminar, Instituto de Matemática, Universidade Federal do Rio de Janeiro

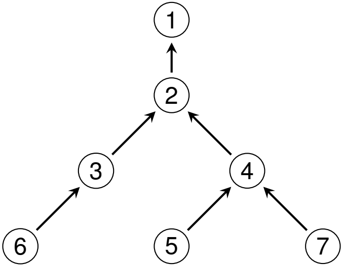
October 16, 2023.

Outline

1. Random cutting (Fortuitous failures)
2. Targeted cutting (Malicious attack)
3. Results
4. Coupling
5. Work in progress

Increasing trees

- Edges directed towards root
- Labels $[n] = \{1, 2, \ldots, n\}$
- Labels increase away from root
- $deg_{T_n}(v)$, $\Delta(T_n)$



Building a random recursive tree

A random recursive tree on n vertices is a tree chosen uniformly at random on the class of increasing trees on $[n]$.

- T_1 is a tree with a single vertex

For $i > 1$ build T_i from T_{i-1} by adding

- vertex i
- arrow $i \to j$

$$\mathbf{P}\left(i \to j | T_{i-1}\right) = \frac{1}{i-1}$$

Building a RRT

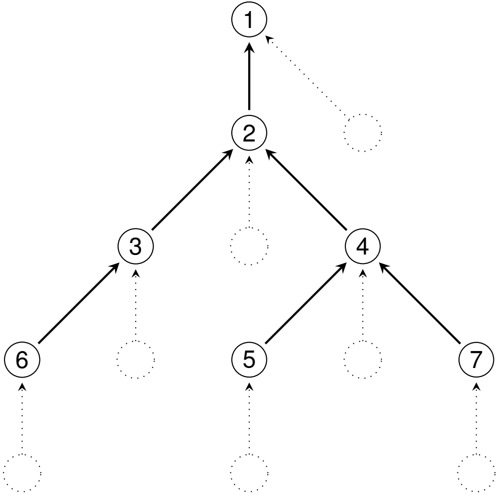


Figure: Obtaining T_8 from T_7

Characteristics

$$D := deg_{T_n}(1) = \sum_{i=2}^{n} B_i, \text{ where } B_i \text{ is a } Bernoulli \left(\frac{1}{i-1}\right)$$

$$\mathbb{E}D = H_{n-1} \sim \ln(n)$$

Goh and Schmutz [2002]. Let $\Delta(T_n)$ be the maximum degree in T_n. Then

$$\Delta(T_n) \sim \log_2(n) \approx 1.4 \ln(n)$$

Characteristics

$$D := deg_{T_n}(1) = \sum_{i=2}^{n} B_i, \text{ where } B_i \text{ is a } Bernoulli\left(\frac{1}{i-1}\right)$$

$$\mathbb{E}D = H_{n-1} \sim \ln(n)$$

Goh and Schmutz [2002]. Let $\Delta(T_n)$ be the maximum degree in T_n. Then

$$\Delta(T_n) \sim \log_2(n) \approx 1.4 \ln(n)$$

Na and Rapoport [1970]

$$Z_i(n) := \#\{1 \leq j \leq n \mid deg_{T_n}(j) = i\},$$

$$\frac{\mathbb{E}Z_i(n)}{n} \xrightarrow{n\to\infty} \frac{1}{2^{i+1}},$$

$$\mathbb{E}Z_i(n) \xrightarrow{n\to\infty} \begin{cases} \infty & \text{for} \quad i << \log_2(n) \\ \frac{1}{2} & \text{for} \quad i - \log_2(n) = \Theta(1) \\ 0 & \text{for} \quad i >> \log_2(n) \end{cases}$$

Characteristics

$$D := deg_{T_n}(1) = \sum_{i=2}^{n} B_i, \text{ where } B_i \text{ is a } Bernoulli\left(\frac{1}{i-1}\right)$$

$$\mathbb{E}D = H_{n-1} \sim \ln(n)$$

Goh and Schmutz [2002]. Let $\Delta(T_n)$ be the maximum degree in T_n. Then

$$\Delta(T_n) \sim \log_2(n) \approx 1.4\ln(n)$$

Na and Rapoport [1970]

$$Z_i(n) := \#\{1 \le j \le n \mid deg_{T_n}(j) = i\},$$

$$\frac{\mathbb{E}Z_i(n)}{n} \xrightarrow{n\to\infty} \frac{1}{2^{i+1}},$$

Addario-Berry and Eslava [2017]

$$Z_{\ge i}(n) = \#\{1 \le j \le n \mid deg_{T_n}(j) \ge i\}.$$

$$\mathbb{E}Z_{\ge k+\lfloor \log_2(n)\rfloor}(n) \xrightarrow{n\to\infty} \frac{1}{2^k}$$

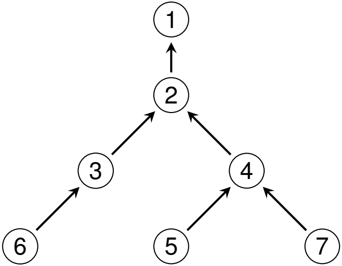
for $k \in \mathbb{N}$ and $k = k(n)$ in a suitable range.

$$\mathbb{E}Z_i(n) \xrightarrow{n\to\infty} \begin{cases} \infty & \text{for} \quad i << \log_2(n) \\ \frac{1}{2} & \text{for} \quad i - \log_2(n) = \Theta(1) \\ 0 & \text{for} \quad i >> \log_2(n) \end{cases}$$

Random cutting

Introduced by Meir and Moon [1974]

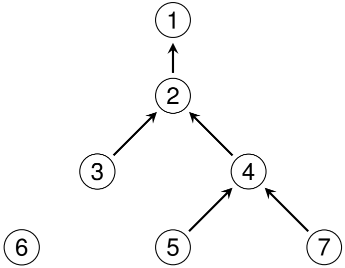
- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

Introduced by Meir and Moon [1974]

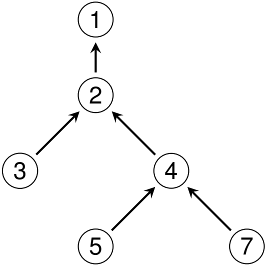
- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

Introduced by Meir and Moon [1974]

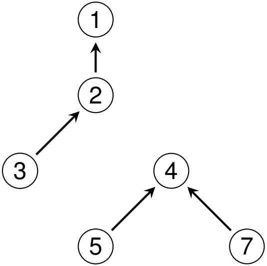
- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

Introduced by Meir and Moon [1974]

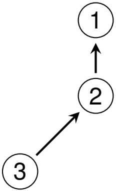
- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

Introduced by Meir and Moon [1974]

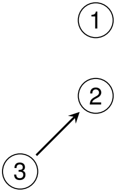
- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

Introduced by Meir and Moon [1974]

- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Random cutting

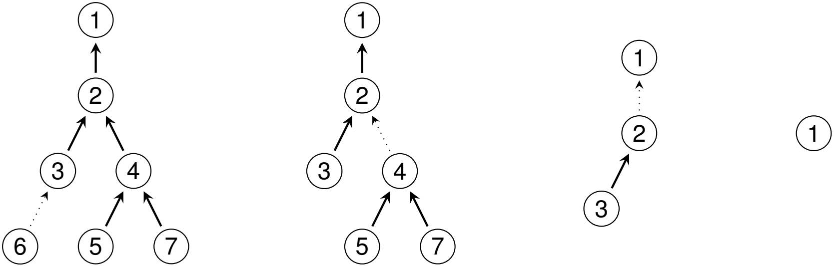
Introduced by Meir and Moon [1974]

- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated

Random cutting

Introduced by Meir and Moon [1974]

- Choose an edge uniformly at random and delete it
- Discard the subtree not containing the root
- Repeat this procedure until the root is isolated



Let $J_n = \#$ cuts needed to destroy a random recursive tree with n vertices.

J_n

Key property: After each cut the remaining tree is a RRT. Used to obtain

$$\mathbb{E}J_n \sim \frac{n}{\ln n}$$

Meir and Moon [1974]

$$\mathbb{E}J_n \sim \frac{n}{\ln(n)}$$

Panholzer [2004]

$$\frac{\ln(n)}{n} J_n \xrightarrow{P} 1$$

Drmota et al. [2009]

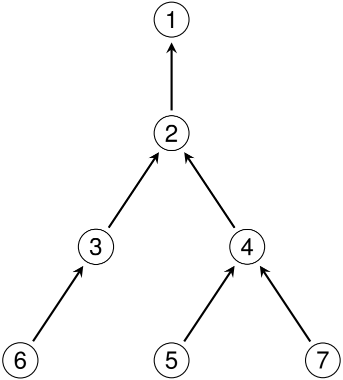
$$\frac{(\ln(n))^2}{n} J_n - \ln(n) - \ln(\ln(n))$$

converges weakly to a random variable Y with characteristic function

$$\varphi_Y(\lambda) = \exp\left\{ i\lambda \ln|\lambda| - \frac{\pi|\lambda|}{2} \right\}$$

Our proposal: Targeted cutting

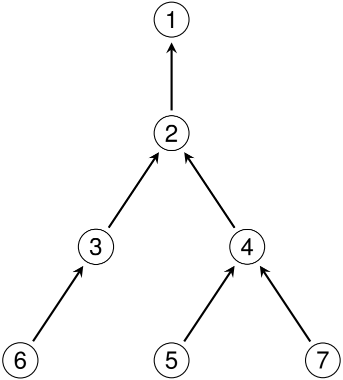
- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed



Our proposal: Targeted cutting

- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed

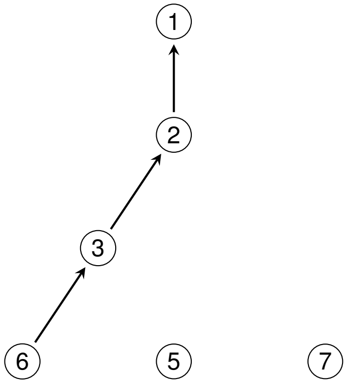
4, 2, 3, 1, 7, 6, 5



Our proposal: Targeted cutting

- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed

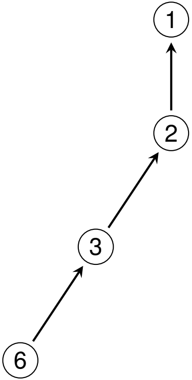
$$4, 2, 3, 1, 7, 6, 5$$



Our proposal: Targeted cutting

- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed

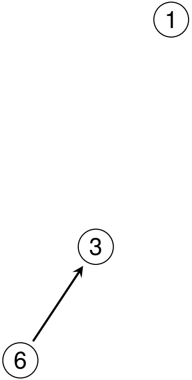
$$4, 2, 3, 1, 7, 6, 5$$



Our proposal: Targeted cutting

- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed

$4, 2, 3, 1, 7, 6, 5$

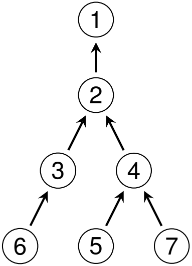


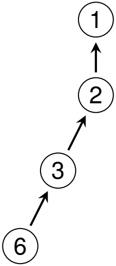
Our proposal: Targeted cutting

- List the vertices from highest to lowest degree
- Sequentially remove the vertex in the listed order, discarding the subtrees not containing the root
- End when the root is removed

$$4, 2, 3, 1, 7, 6, 5$$

Our proposal: Targeted cutting





4, 2, 3, 1, 7, 6, 5

Our proposal: Targeted cutting

We lose properties that were known in the random cutting procedure, including

- The distribution of the remaining tree
- The number of vertices deleted after each cut

Let $K_n = \#$ targeted cuts needed to delete a random recursive tree with n vertices.
Recall

- $Z_{\geq i}(n) = \#\{1 \leq j \leq n \mid deg_{T_n}(j) \geq i\}$
- $D = deg_{T_n}(1)$

Note

$$K_n \leq Z_{\geq D}(n).$$

Results

Let $\gamma := 1 - \ln(2) \approx 0.307$

Theorem (Moments)

For any positive integer k,

$$\mathbb{E}\big(\ln(Z_{\geq D})\big)^k = \big(\ln(n^\gamma)\big)^k (1 + O(1)).$$

Theorem (Convergence in probability)

$$\frac{\ln(Z_{\geq D})}{\ln(n)} \xrightarrow{P} \gamma.$$

Theorem (Growth order)

For any $\varepsilon > 0$,

$$K_n = O_p(n^{\gamma+\varepsilon}).$$

Proof strategy

Conditioned on the event

$$(1-\varepsilon)\ln(n) \leq D \leq (1+\varepsilon)\ln(n),$$

which we denote $D \in (1 \pm \varepsilon)\ln(n)$, we have that

$$Z_{\geq \lceil (1+\varepsilon)\ln(n)+1 \rceil} \leq Z_{\geq D} \leq Z_{\geq \lfloor (1-\varepsilon)\ln(n) \rfloor}.$$

Moments of the upper bound can be controlled [Addario-Berry and Eslava, 2017].
Lower bound can be approximated with a Poisson random variable [Eslava, 2020].

Coupling

The idea is to build simultaneously T_n, a RRT, and $T_n^{(\varepsilon)}$, a RRT conditioned on $D^{(\varepsilon)} \in (1 \pm \varepsilon)\ln(n)$.

- $B = (B_2, \ldots, B_n)$ with indepent entries where B_i is Bernoulli $\left(\frac{1}{i-1}\right)$.
- $B^{(\varepsilon)} = (B_2^{(\varepsilon)}, \ldots, B_n^{(\varepsilon)})$ with the law of B conditioned on $\sum_{i=2}^n B_i^{(\varepsilon)} \in (1 \pm \varepsilon)\ln(n)$.
- Y_i uniform in $\{2, 3, \ldots, i-1\}$ for $i \geq 3$.

Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$ $\quad$ $B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$

Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

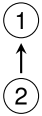
$B = (1, 0, 0, 0, 0, 0)$

$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$

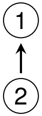
Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$$B = (1, 0, 0, 0, 0, 0)$$



$$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$$



Coupling

$$Y_3 = 2,\, Y_4 = 2,\, Y_5 = 4,\, Y_6 = 3,\, Y_7 = 4$$

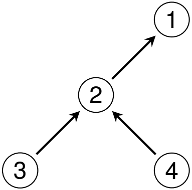
$B = (1, 0, 0, 0, 0, 0)$

$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$

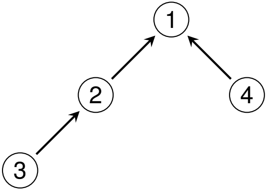
Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$



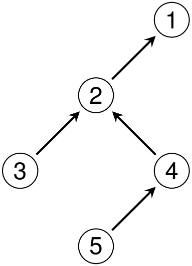
$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$



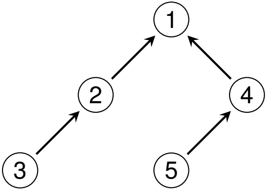
Coupling

$$Y_3 = 2,\, Y_4 = 2,\, Y_5 = 4,\, Y_6 = 3,\, Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$



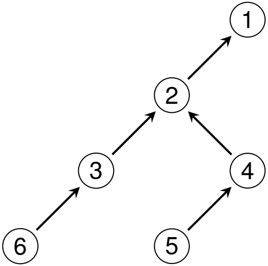
$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$



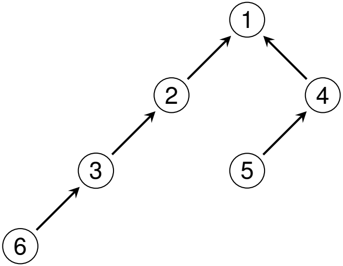
Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$



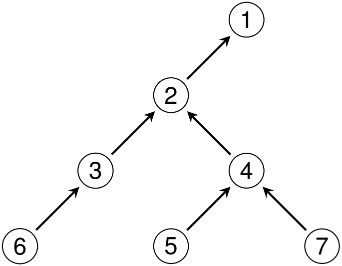
$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$



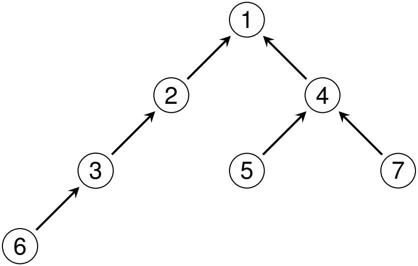
Coupling

$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$



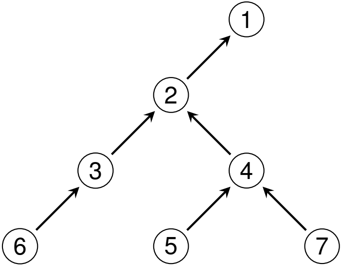
$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$



Coupling

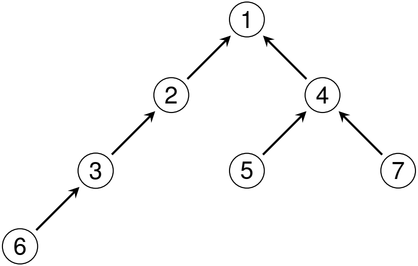
$$Y_3 = 2,\ Y_4 = 2,\ Y_5 = 4,\ Y_6 = 3,\ Y_7 = 4$$

$B = (1, 0, 0, 0, 0, 0)$



T_n

$B^{(\varepsilon)} = (1, 0, 1, 0, 0, 0)$



$T_n^{(\varepsilon)}$

Under this coupling

Let $Z^{(\varepsilon)}_{\geq i}(n) = \#\{1 \leq j \leq n \mid deg_{T_n^{(\varepsilon)}}(j) \geq i\}$, $W_d = \dfrac{1 + Z^{(\varepsilon)}_{\geq d}(n)}{1 + Z_{\geq d}(n)}$

Proposition (Similitud en razón)

Let $\varepsilon \in (0, 1/3)$, $\delta \in (0, 1)$. *There exists* $\beta_3 \in (0, \varepsilon^2/12)$, *C constant y* $n_0 = n_0(\varepsilon, \delta)$ *such that* $0 < d < (1+\varepsilon)\ln(n)$ *under the coupling*

$$\mathbb{P}\left(W_d \in (1 \pm \delta)\right) \geq 1 - Cn^{-\beta_3}.$$

for each $n > n_0$

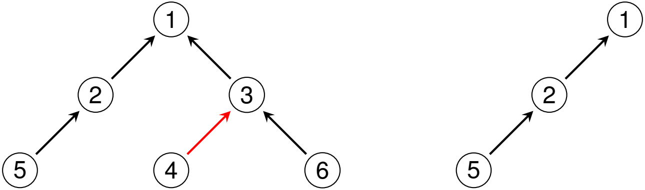
Proposition

Let $\varepsilon \in (0, 1/3)$, $k \in \mathbb{N}$. *For* $0 < d < (1+\varepsilon)\ln(n)$ *under the coupling*

$$\left|\mathbb{E}(\ln(W_d))^k\right| \leq 2C(\ln(n+1))^k n^{-\beta_3}.$$

Work in progress: Degree biased random cutting

- Choose an edge uniformly at random and delete the vertex closest to the root.
- Keep the subtree having the root.
- The process is finished when the root is deleted.



$$\frac{(\ln n)^2}{n}\mathbb{E}(Z_n) \to 1.$$

Open problems

- $\delta-$atack: Generalized cutting process. Let $\delta \in (0,1)$ and T_n a RRT.
  - List the vertices by their degree.
  - Remove a vertex uniformly from $\{v_1, \dots, v_{\lceil \delta n \rceil}\}$.
  - End the process when the root is removed.

  Case $\delta \geq 1 - \frac{1}{n}$ is the random atack, case $\delta \leq \frac{1}{n}$ is targeted cutting. Interesting case: $\delta = n^{-\alpha}$ for some $\alpha > 0$.
- Weighted random trees.

References

L. Addario-Berry and L. Eslava. High degrees in random recursive trees. *Random Structures and Algorithms*, 52(4):560–575, 2017.

M. Drmota, A. Iksanov, M. Möhle, and U. Roesler. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures and Algorithms*, 34(3):319–336, 2009.

L. Eslava. A non-increasing tree growth process for recursive trees and applications. *Combinatorics, Probability and Computing*, pages 1–26, 2020.

W. Goh and E. Schmutz. Limit distribution for the maximum degree of a random recursive tree. *Journal of Computational and Applied Mathematics*, 142(1):61–82, 2002. Probabilistic Methods in Combinatorics and Combinatorial Optimization.

A. Meir and J. Moon. Cutting down recursive trees. *Bellman Prize in Mathematical Biosciences*, 21:173–181, 1974.

H.S. Na and A. Rapoport. Distribution of nodes of a tree by degree. *Bellman Prize in Mathematical Biosciences*, 6:313–329, 1970.

A. Panholzer. Destruction of recursive trees. *Mathematics and Computer Science III*, 528:267–280, 2004.

Recapitulation

- Number of random cuts needed to destroy a RRT of size n:

$$\mathbb{E}J_n \sim \frac{n}{\ln(n)}.$$

$$\frac{\ln(n)}{n} J_n \xrightarrow{P} 1.$$

$$\frac{(\ln(n))^2}{n} J_n - \ln(n) - \ln(\ln(n)) \xrightarrow{d} Y.$$

- Number of targeted cuts needed to destroy a RRT K_n:
  For any $\varepsilon > 0$,

$$K_n = O_p\left(n^{1-\ln(2)+\varepsilon}\right).$$

- $K_n \leq Z_{\geq D}(n)$:

$$\frac{\ln(Z_{\geq D}(n))}{\ln(n)} \xrightarrow{P} 1 - \ln(2).$$

$$\mathbb{E}\bigg(\ln(Z_{\geq D}(n))\bigg)^k \sim \left(\ln\left(n^{1-\ln(2)}\right)\right)^k.$$