Probabilistic Aspects of Computer Science
Random sampling

Anne Bouillard

November 22, 2020
Approximation of $\pi$: the Monte-Carlo method

Simulation, or random generation, is a set of techniques aiming at estimating a quantity or sampling according to a probability distribution, usually when these cannot be efficiently computed.

- Consider a square with size of length 2 and its incircle
- the area of the circle is $\pi$ and that of the square is 4.

- Choose a point $(X, Y)$ uniformly in the square.
- let
  
  $$Z = \begin{cases} 
  1 & \text{if } \sqrt{X^2 + Y^4} \leq 1 \quad \text{(point in the circle)}, \\
  0 & \text{otherwise.}
  \end{cases}$$

- $P(Z = 1) = \pi/4$. 
Approximation of $\pi$: the Monte-Carlo method (2)

- Choose $m$ uniformly distributed and independent points of the square $(X_i, Y_i)_{1 \leq i \leq m}$.
- one can associate $Z_i$ to each $(X_i, Y_i)$
- Set $W = \sum_{i=1}^{m} Z_i$, then

$$E[W] = E[\sum_{i=1}^{m} Z_i] = \sum_{i=1}^{m} E[Z_i] = \frac{m\pi}{4}.$$  

$W' = \frac{4}{m} W$ is an estimator of $\pi$. Using Chernoff bounds:

$$P(|W' - \pi| \geq \epsilon\pi) \leq P(|W - \frac{m\pi}{4}| \geq \epsilon\frac{m\pi}{4}) \leq 2e^{-\frac{m\pi\epsilon^2}{32}} = 2e^{-\frac{m\pi\epsilon^2}{12}}.$$

**Definition**

A probabilistic algorithm returns an $(\epsilon, \delta)$-approximation for $v$ if the output $X$ of the algorithm satisfies

$$P(|X - v| \leq \epsilon v) \geq 1 - \delta.$$
Approximation of $\pi$: the Monte-Carlo method (3)

In our case, when $\epsilon < 1$,

$$2e^{\frac{-m\pi\epsilon^2}{12}} \leq \delta \iff \ln 2 - \frac{m\pi\epsilon^2}{12} \leq 1 \ln \delta \iff \frac{\epsilon^2m\pi}{12} \geq \ln \frac{2}{\delta} \iff m \geq \frac{12}{\pi\epsilon^2} \ln 2\delta.$$

We have an $(\epsilon, \delta)$-approximation for computing $\pi$ consisting in computing $W'$ for $m \geq \frac{12}{\pi\epsilon^2} \ln 2\delta$.

**Definition (FPRAS)**

A fully polynomial randomized approximation scheme for a problem returns for each entry $x$ and all $\delta, \epsilon > 0$ an $(\epsilon, \delta)$-approximation of $v(x)$ in time polynomial in $\frac{1}{\epsilon}$, $\ln \frac{1}{\delta}$ and the size of $x$.

Here the complexity is linear on $m$, so we have a polynomial scheme.
From a uniform distribution on $[0, 1]$ to discrete distributions

Let $\pi$ be a probability distribution on $\mathbb{N}$ or $\{1, \ldots, n\}$.

**Algorithm 1:** Sample according to distribution $\pi$

\begin{verbatim}
begin
    ~ Uniform([0, 1]);
i ← 0; S ← π(0);
while U > S do
    i ← i + 1;
    S ← S + π(i)
Return i.
\end{verbatim}

The returned value $i$ satisfies

$$\sum_{k=1}^{i-1} \pi(k) < U \leq \sum_{k=1}^{i} \pi(k).$$

The probability to return $i$ is then exactly $\pi(i)$.

This method can be applied for continuous distributions and is then called the inverse method.
The rejection sampling

**Idea:** consider the example of the circle, and suppose we want to sample a point uniformly in the circle. It might not be very easy to draw it uniformly. One way is to draw an angle and a radius. But the radius cannot be uniformly sampled. An easier way is to sample a point in the square until the point is in the circle.

Note: the method is also valid for continuous probabilities.

- Let \((p_i)_{i \in \mathbb{N}}\) be a probability distribution
- and \((s_i)_{i \in \mathbb{N}}\) be a sequence of positive numbers, with \(\sum_{i \in \mathbb{N}} s_i = S < \infty\) and
  - for all \(i\), \(p_i \leq s_i\).
- \((s_i)\) can be normalized to define a probability distribution: \((q_i) = \left(\frac{s_i}{S}\right)\)

We assume that \((q_i)\) is easy to sample, and the goal is to sample \((p_i)\).
The rejection sampling (2)

Algorithm 2: Rejection sampling

begin
    while $U_n < \frac{p_{Y_n}}{s_{Y_n}}$ do
        Sample $Y_n$ according to $(q_i)$;
        Sample $U_n \sim \text{Unif}([0, 1])$
    return $Y_n$

$$P(U_n \leq \frac{p_{Y_n}}{s_{Y_n}}) = \sum_{i \in \mathbb{N}} P(U_n \leq \frac{p_{Y_n}}{s_{Y_n}} \mid Y_n = i) P(Y_n = i) = \frac{p_i}{s_i} s = \sum_{i \in \mathbb{N}} \frac{p_i}{S} = \frac{1}{S},$$

Let $N$ be the number of loops executed. This is a random variable and the algorithm returns $Y_N$.

Theorem

$Y_N$ is distributed according to $(p_i)$. 
The rejection sampling - Proof

First compute $\mathbf{P}(X_n = i, N = n)$:

$$\mathbf{P}(Y_n = i, N = n) = \mathbf{P}(U_1 < \frac{pY_1}{sY_1}, \ldots, U_{n-1} < \frac{pY_{n-1}}{sY_{n-1}}, U_n \geq \frac{pY_n}{sY_n}, Y_n = i)$$

$$= (1 - \frac{1}{S})^{n-1}\mathbf{P}(U_n \geq \frac{pY_n}{sY_n}, Y_n = i)$$

$$= (1 - \frac{1}{S})^{n-1}\frac{p_i}{S}.$$

We then deduce

$$\mathbf{P}(Y_n = i) = \sum_{n \in \mathbb{N}} \mathbf{P}(Y_n = i, N = n) = \sum_{n \in \mathbb{N}} (1 - \frac{1}{S})^{n-1}\frac{p_i}{S} = p_i.$$  

This method is efficient when $\mathbb{E}[N]$ is small.
The aliasing method - example

**Goal:** Sampling on a finite space $E = \{0, \ldots, n\}$ a large number of times, which can become costly. Here the sampling time, after some pre-computation with complexity $O(n)$ is constant for each sampling: it is obtained by sampling two uniform variables, one on $\{1, \ldots, n\}$ and the other on $[0, 1]$

Take $n = 4$, $\pi(1) = \frac{3}{10}$, $\pi(2) = \frac{4}{10}$, $\pi(3) = \frac{2}{10}$ and $\pi(4) = \frac{1}{10}$.
The aliasing method - algorithm

Algorithm 3: Aliasing method

begin

for $i \in \{1, \ldots, n\}$ do

$q_i \leftarrow np_i; s_i \leftarrow 0; h_i \leftarrow 0$;

$G \leftarrow \{i \mid q_i > 1\}$;

$H \leftarrow \{i \mid q_i < 1\}$;

while $H \neq \emptyset$ do

choose $j \in H$ and $k \in G$;

$s_j \leftarrow 1 - q_j$;

$q_k \leftarrow q_k + q_j - 1$;

$h_j \leftarrow k$;

$H \leftarrow H \setminus \{j\}$;

if $q_k < 1$ then $H \leftarrow H \cup \{k\}$;

$G \leftarrow G \setminus \{k\}$;

return $\{(s_j, h_j), j \in E\}$

end

To sample, it is now enough to draw $j$ uniformly on $\{1, \ldots, n\}$, and $p$ uniformly on $[0, 1]$. If $p < s_j$, return $h_j$ else, return $j$. 
Monte Carlo Markov Chains

Sampling according to the stationary distribution of a Markov chain

For an ergodic Markov chain:

1. compute the stationary distribution $\pi$;
2. sampler $\pi$ according to the previous algorithms.

Problems:

- $\pi$ can be difficult to compute, if the state-space is too large;
- even the state space can be too complex to be computed.

We will then see specific means to sample Markov chains.
Example: independent sets

- $G = (V, E)$ an undirected graph with vertices $V = \{1, \ldots, n\}$ and edges $E$.
- an independent set of $G$ is a set of vertices $I$ that are not adjacent to each other $I \subseteq V$ is an independent set iif $\forall u, v \in I, u \neq v, \{u, v\} \notin A$.

**Goal:** sample an independent set according to the uniform distribution.

1. **naive method:** generate all the independent set. This is impossible in practice: for a graph with $n$ nodes, there are $2^n$ subsets to check (even if smarter method may exist).
2. **rejection sampling:** the problem is that the number of independent sets can be small compared to the number of subsets of vertices

### Gibbs sampling

1. Construct an ergodic Markov chain whose stationary distribution is then uniform distribution of independent sets.
2. Simulate this Markov chain long enough, so that the state obtained is almost distributed according to this stationary distribution (Kolmogorov theorem).
Gibbs sampler and product-form distribution

**Idea:** Technique to build a reversible Markov chain whose distribution has a product-form:

- State space $\mathcal{E} \subseteq Q^n$,
- there exists $Z$ a normalizing constant such that

$$\pi(q_1, \ldots, q_n) = \frac{1}{Z} \pi_1(q_1) \times \cdots \times \pi_n(q_n),$$

and $\pi_i$ is a probability distribution on $Q$.

### Product-form distribution examples

Consider $Q = \{0, 1\}$,

1. **Uniform distribution:** $\pi_i(0) = \pi_i(1) = 1/2$, $\pi(q_1, \ldots, q_n) = \frac{2^{-n}}{Z} = |E|^{-1}$ so $Z = |E|2^n$.

2. **Boosting "large states":** $\pi_i(k) = \frac{\lambda^{|k|}}{1+\lambda}$. So

$$\pi(q_1, \ldots, q_n) = \frac{1}{Z} \frac{\lambda^{\sum q_i}}{(1 + \lambda)^n}.$$
Product-form distribution and reversible Markov chains

- **Target distribution**
  \[ \pi(q_1, \ldots, q_n) = \frac{1}{Z} \pi_1(q_1) \times \cdots \times \pi_n(q_n), \]

- **Constraint on the transition matrix**: \(P\) such that \(x, y \in \mathbb{Q}^n, P(x, y) \neq 0\) only if \(x\) and \(y\) differ at most for one coordinate.
  - suppose \(x_1 \neq y_1\) and \(x_i = y_i\) for all \(i \neq 1\).
  - We choose \(p_{x,y} = \alpha \pi_1(y_1)\). So there exists a constant \(\alpha\) such that
    \[ \pi_1(x_1)p_{x,y} = \alpha \pi_1(x_1)\pi_1(y_1) = \pi_1(y_1)p_{y,x}. \]

- **The Markov chain is reversible**:
  \[ \pi(x)p_{x,y} = \frac{\pi_1(x_1) \cdots \pi_n(x_n)}{Z} \alpha \pi_1(y_1) = \pi(y)p_{y,x}. \]

The constant \(\alpha\) corresponds to the choice of a distribution to choose the coordinate the transition is made.
Gibbs sampler in the general case

The algorithm describes one step of the sampling: from state $x = (x_1, \ldots, x_n)$,

1. Choose the coordinate $i$ to modify
2. compute $y_i$ according to $\pi(i)$. Then new state can depend on $x$, except on coordinate $i$, as the new state must belong to $\mathcal{E}$.

Algorithm 4: Gibbs sampler

Input: $x \in Q^S$

begin
  Draw $i \in \{1, \ldots, n\}$ according to distribution $d$;
  foreach $j \neq i$ do $y_j \leftarrow x_j$;
  Draw $k \in Q$ according to distribution $\pi_i$ and independently of $d$;
  if $x_1, \ldots, y_i, \ldots, x_n \in \mathcal{E}$ then
    $y_i \leftarrow k$;
  else
    $y_i \leftarrow x_i$;
  Return $y$.  

Gibbs sampler for the independent sets

Algorithm 5: G.S. for the independent sets

Input: \( I \subseteq V \)

begin

Draw \( u \in V \) according to the unif. distr.;

if \( \forall v \) neibourgh of \( u \), \( v \notin I \) then

\[ I' \leftarrow I \cup \{ u \} \text{ with probability } 1/2; \]

\[ I' \leftarrow I \setminus \{ u \} \text{ with probability } 1/2; \]

else

\[ I' \leftarrow I \]

Return \( I' \).

Let \( I_1 \) and \( I_2 \) be two independent sets, that differ only for one vertex:

\[ I_1 = I_2 \cup \{ u \}. \]

Then

\[ P(x, y) = \frac{1}{n} \times \frac{1}{2} = P(y, x), \]

so the stationary distribution is the uniform distribution.

The Markov chain is ergodic:

- each independent set \( I \) is reachable from the empty set: one can add the vertices of \( I \) one by one (possible with probability 1/2 at each step)
- Similarly, the empty set is reachable from all independent sets.
- For all \( x \in I \), \( P(x, x) \geq 1/2 \) for all \( x \), so the Markov chain is aperiodic.
MCMC (Monte-Carlo Markov chain) simulation

The construction of the Markov chain is done. It is time to simulate:

- From state $x$;
- perform $r$ steps of the simulation. Then $X_r$ is a sample whose distribution is almost $\pi$;
- perform $r$ more steps. $X_r$ is another sample whose distribution is almost $\pi$ and almost independent of $X_r$;
- ...

What does *almost* mean?

For example, to estimate the average size of an independent set, the error is

$$\sum_{I} |I| (P(X_r = I) - \pi(I)) \leq \sum_{I} |I| |P(X_r = I) - \pi(I)| \leq n \sum_{I} |P(X_r = I) - \pi(I)|.$$  

We would like to control this error $\sum_{I} |P(X_r = I) - \pi(I)| \leq \epsilon$ for a small $\epsilon$. For this we have to choose $r$ large enough.
Distance in total variation

Definition

Let $\mu_1$ and $\mu_2$ two probability distributions on $E$ (at most denumerable). The Distance in (total) variation between $\mu_1$ and $\mu_2$ is

$$||\mu_1 - \mu_2||_{TV} = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|.$$  

- $0 \leq ||\mu_1 - \mu_2||_{TV} \leq 1$
- $||\mu_1 - \mu_2||_{TV} = 0$ if $\mu_1 = \mu_2$
- $||\mu_1 - \mu_2||_{TV} = 1$ if $\mu_1$ and $\mu_2$ have disjoint supports.
Distance in total variation - Example

\[ ||\mu_1 - \mu_2||_{TV} = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|. \]

Example

uniform distribution \( \mu_1 \) et la distribution \( \mu_2 = (1/2, 1/4, 1/8, 1/8). \)

\[ ||\mu_1 - \mu_2||_{TV} = \frac{1}{2}(0, 25 + 2 \times 0, 125) = 0, 25. \]
Distance in total variation - Equivalent definition

Lemma (Equivalent definition)

For all $A \subseteq E$, set $\mu_i(A) = \sum_{x \in A} \mu_i(x)$. Then

$$||\mu_1 - \mu_2||_{TV} = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$  

- $E^+$ the set of elements of $E$ such that $\mu_1(x) \geq \mu_2(x)$
- $E^-$ the set of elements of $E$ such that $\mu_1(x) < \mu_2(x)$.
- We have $\mu_1(E^+) - \mu_2(E^+) = \max_{A \subseteq E} \mu_1(A) - \mu_2(A)$
- But $\mu_1(E^+) - \mu_2(E^+) = \mu_2(E^-) - \mu_1(E^-) \geq 0$
- So

$$|\mu_1(E^+) - \mu_2(E^+)| + |\mu_1(E^-) - \mu_2(E^-)| = \sum_{x \in E} |\mu_1(x) - \mu_2(x)| = 2||\mu_1 - \mu_2||_{TV}.$$
Example: shuffling a card deck

- 52 cards in the deck
- a card is chosen uniformly at random and placed at the top of the deck
- one can represent this process by a Markov chain on the possible orders of the cards in the deck
- the stationary distribution is the uniform distribution:

\[
\pi_x = \frac{1}{52} \sum_{y \in N_x} \pi_y,
\]

where \( N_x \) is the set of successors of \( x \) : \(|N_x| = 52\).

- If \( \mu_1 \) is the uniform distribution and
- \( \mu_2 \) the uniform distribution on the order having the ace of spades at the top,

\[
||\mu_1 - \mu_2|| \geq 1 - \frac{1}{52} = \frac{51}{52}.
\]

Indeed, we take \( A \) as the set of decks having the ace of spades at the top.
Mixing time of a Markov chain

- $\pi$ is the stationary distribution of a Markov chain $\{X_n\}$ on state space $E$.
- $p^n_x$ is the distribution of $X_n$ if the initial state is $x$.

$$\Delta_x(n) = ||p^n_x - \pi|| \quad \text{and} \quad \Delta(n) = \max_x \Delta_x(n),$$

$$\tau_x(\epsilon) = \min\{n \mid \Delta_x(n) \leq \epsilon\} \quad \text{and} \quad \tau(\epsilon) = \max_{x \in E} \tau_x(\epsilon).$$

- $\tau(\epsilon)$ is called the **mixing time** of the Markov chain
- A Markov chain is **rapidly mixing** if $\tau(\epsilon)$ is polynomial in $\log(\frac{1}{\epsilon})$ and in the size of the problem.
- **Warning:** the size of the problem ($52 \log_2 52$) $\neq$ size of the state space ($52!$)
The spectral method (finite state space)

**Idea:** use the results of linear algebra on primitive matrices:
The transition matrix of an ergodic Markov chain (irreducible and aperiodic) is a stochastic and primitive matrix.

**Theorem**

*Perron-Frobenius* If \( P \) stochastic and primitive:

- **1 is the largest eigenvalue (in modulus) associated to** \( P \)
- **\( P \) a unique unitary left-eigenvector** \( \pi \) **associated to** 1.
  - \( \pi \) **is positive**
  - **this proves the existence and uniqueness of the stationary distribution**
- **\( P^n \to 1\pi \)**
- **if \( \lambda \) is the modulus of the second largest eigenvalue of** \( P \)**, then with** \( \pi_{min} = \min_{j \in \mathcal{E}} \pi(j) \),

\[
||\mu p^n - \pi||_{TV} \leq \frac{1}{2\pi_{min}} \lambda^n.
\]
The spectral method (2)

\[ ||(\mu p^n) - \pi||_{TV} \leq \frac{1}{2\pi_{\text{min}}} \lambda^n. \]

Constant \(\frac{1}{2\pi_{\text{min}}}\) can be improved but the bounds found with this method are usually not precise enough:

\[ \epsilon \geq \frac{\lambda^n}{2\pi_{\text{min}} \pi(j)} \Rightarrow n \geq \frac{\ln(2\pi_{\text{min}} \epsilon)}{\ln \lambda}. \]

We will use the coupling of Markov chains.
Coupling of a Markov chain

General technique to bound the mixing time

Definition

The coupling of a Markov chain \((X_n)_{n \in \mathbb{N}}\) on state space \(\mathcal{E}\) is a Markov chain \(Z_n = (X^1_n, X^2_n)\) on \(\mathcal{E} \times \mathcal{E}\) such that

\[
\begin{align*}
P(X^1_{n+1} = x' \mid Z_n = (x, y)) &= P(X_{n+1} = x' \mid X_n = x) \\
P(X^2_{n+1} = y' \mid Z_n = (x, y)) &= P(X_{n+1} = y' \mid X_n = y).
\end{align*}
\]

1. \((X^1_n)\) and \((X^2_n)\) transition matrix as \((X_n)\)
2. \((X^1_n)\) and \((X^2_n)\) are not necessarily independent.

We are interested in coupling that

- make \((X^1_n)\) and \((X^2_n)\) rejoin \((\exists n \text{ such that } X^1_n = X^2_n)\);
- Once they rejoin, the remain equal.
Coupling and mixing time

**Idea:** choose coupling that make chains join as quick as possible.

For example, if $f$ is a functional representation of $X_n$ $X_{n+1} = f(X_n, U_{n+1})$, one can choose the representation

$$(X_{n+1}^1, X_{n+1}^2) = (f(X_n^1, U_{n+1}), f(X_n^2, U_{n+1})).$$

That way, once $X^1$ and $X^2$ rejoin, they remain the same.

**Theorem**

Let $Z_n = (X_n, Y_n)$ be a coupling of a Markov chain on $\mathcal{E}$. Suppose there exists $T$ such that $\forall x, y \in \mathcal{E}$, $P(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \epsilon$. Then $\tau(\epsilon) \leq T$. 
Coupling and mixing time - proof

1. Choose the coupling where $Y_0 = \pi$ and $X_0 = x \in E$.

\[ P(X_T \neq Y_T) = \sum_y P(X_T \neq Y_T \mid Y_0 = y)\pi(y) \leq \sum_y \epsilon \pi(y) \leq \epsilon. \]

2. Let $A \subseteq E$. We first bound $|P(X_T \in A) - P(Y_T \in A)|$.

\[
P(X_T \in A) \geq P(X_T = Y_T \text{ et } Y_T \in A) \\
\geq 1 - P(X_T \neq Y_T \text{ ou } Y_T \notin A) \\
\geq 1 - P(X_T \neq Y_T) - P(Y_T \notin A) \\
\geq P(Y_T \in A) - \epsilon = \pi(A) - \epsilon.
\]

3. Similarly when $A$ is replaced by $E \setminus A$ we get $P(X_T \notin A) \geq p(Y_T \notin A) - \epsilon$ and $P(X_T \in A) \leq P(Y_T \in A) + \epsilon = \pi(A) + \epsilon$. so

\[
\max_A |p^nx(A) - \pi(A)| \leq \epsilon.
\]
Slow coupling in card shuffling

1. Independent evolutions
   - The coupling is almost surely finite
   - After joining, the Markov chains still have independent evolutions...

2. Choose a position \( j \) uniformly at random, put the card at this position on the top.
   - If there is coupling, the chains remain the same
   - If the initial distribution are different, the chains never join...
Rapid coupling in card shuffling

1. First chain: choose a position $j$ at random, put the card $c$ at the $j$-th position on the top

2. Second chain: put card $c$ at the top. As the position $j$ is chosen uniformly at random, so is $c$.

- A card put at the top will remain at the same position in the two decks in the future evolution.
- The coupling happens when all cards (except one) have been chosen at least once
- From the coupon collector: the probability that a card has not been chosen after $n \ln n + cn$ steps is

$$n(1 - \frac{1}{n})^{n \ln n + cn} \leq e^{\ln n + c} = \frac{e^{-c}}{n}.$$ 

- The probability that there exists a card not chosen is then $e^{-c}$, and

$$\tau(\epsilon) \leq n \ln n + n \ln \frac{1}{\epsilon}.$$
Card shuffling and *cut-off* phenomenon

- The chain is reversible: the reversed chain is the same, but consist in taking the top card and inserting at random in the deck.

- One wants to find a (stopping) time $T$ such that at $T$, the packet is shuffled: for all permutation $\sigma$ of the cards, $P(X_k = \sigma \mid T = k) = \frac{1}{n!}$.

- This time is obtained when the card initially at the bottom is at the top and then inserted.
  - let bottom be the card initially at the bottom of the deck
  - $C_n$ is the set of cards below bottom at step $n$
  - This holds for $C_0 = \emptyset$
  - if this holds for $C_n$, then it holds for $C_{n+1}$: either $C_n = C_{n+1}$, or a card is inserted uniformly in $C_n$.

- The same argument as previously leads to $P(T > n \ln n + cn) \leq e^{-c}$. 

Card shuffling and cut-off phenomenon: lower bound on the mixing time

- Set $k = k(n) = n \ln n + c_n n$, with $c_n \to \infty$.
- $T_j$ step when the $j$-t card is inserted below bottom for the first time.
- $T_{i+1} - T_i$ is the time needed for one more card places under bottom. So $T_{i+1} - T_i \sim \text{Geo}\left(\frac{i+1}{n}\right)$ and $\mathbb{E}[T_{i+1} - T_i] = \frac{n}{i+1}$ and $\text{Var}(T_{i+1} - T_i) \leq \frac{n^2}{(i+1)^2}$.
- Let $x$ be the initial deck and $A_j$ the set of decks where the relative order on the $j$ last cards is the same as $x$. Then $\mathbb{P}(X_k \in A_j) \geq \mathbb{P}(T - T_{j-1} > k)$.
- As $T - T_j = \sum_{i=j}^{n-1} T_{i+1} - T_i$ and the $T_{i+1} - T_i$ and mutually independent,
  $$\mathbb{E}[T - T_j] = \sum_{i=j}^{n-1} \frac{n}{i+1} = n \log n + O(n) \quad \text{ and } \quad \text{Var}(T - T_j) \leq \sum_{i=j}^{n-1} \frac{n^2}{(i+1)^2} = O(n^2).$$
- From the Tchebychev inequality, $\mathbb{P}(T - T_{j-1} \leq k) \leq \frac{\text{Var}(T-T_{j-1})}{n^2 c_n^2} \xrightarrow{n \to \infty} 0$. 