Quickselect under Yaroslavskiy's Dual-Pivoting Algorithm

Sebastian Wild   Markus E. Nebel   Hosam Mahmoud

wild@cs.uni-kl.de

Computer Science Department

Technische Universität Kaiserslautern

28 May 2013

AofA 2013

Menorca

Selection

Consider the following problem

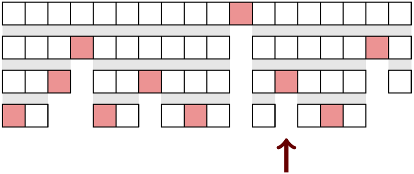
The Selection Problem

Given an **unsorted** array `A` of n data elements,
find the **m-th smallest** element.

- **Applications**
  - estimate distribution of data collection in statistics
  - world rankings in sports
- **Algorithms**
  - **trivial solution:** `sort(A); return A[m];` $\mathcal{O}(n \log n)$
  - linear worst case possible: *Median-of-Medians* (Blum et al.) but rather slow in practice (compared to sorting!)
  - Hoare 1961: **Quicksort** idea usable for selection

From Quicksort to Quickselect

Sort whole array by Quicksort

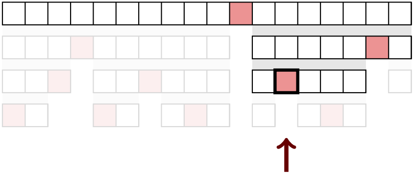


- Markus Nebel just showed:
  Yaroslavskiy's partitioning can **speed up Quicksort**

Does success of dual partitioning in sorting carry over to selection?

From Quicksort to Quickselect

Can prune many branches

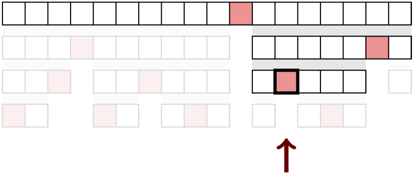


- Markus Nebel just showed:
  Yaroslavskiy's partitioning can **speed up Quicksort**

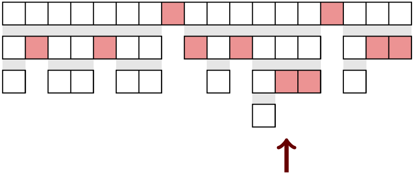
Does success of dual partitioning in sorting carry over to selection?

From Quicksort to Quickselect

Can prune many branches



Dual Pivot Quicksort

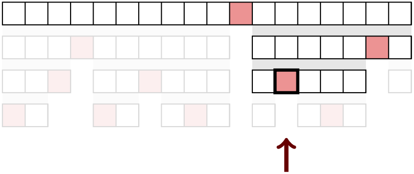


- Markus Nebel just showed:
  Yaroslavskiy's partitioning can **speed up Quicksort**

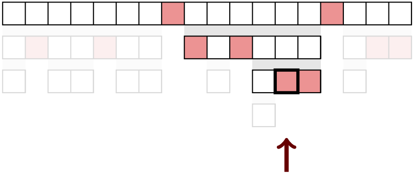
Does success of dual partitioning in sorting carry over to selection?

From Quicksort to Quickselect

Can prune many branches



Prune even more branches



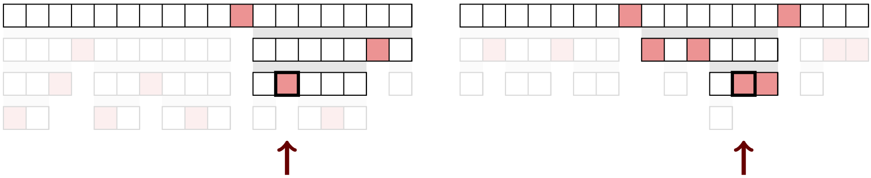
- Markus Nebel just showed:
  Yaroslavskiy's partitioning can **speed up** **Quicksort**

Does success of dual partitioning in sorting carry over to selection?

From Quicksort to Quickselect

Can prune many branches

Prune even more branches



- Markus Nebel just showed:
  Yaroslavskiy's partitioning can speed up **Quicksort**

Does success of dual partitioning in sorting carry over to selection?

Notation

- Analyse **(random)** number of data **comparisons $C_n^{(m)}$** to select from random permutation

- **P** < **Q**: (random) **ranks** of the two pivots

⇝ **subarray sizes** determined by P and Q:

left	P	middle	Q	right
$J_1 := P - 1$		$J_2 := Q - P - 1$		$J_3 := n - Q$

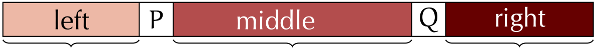
- selection continues recursively in — corresponding **events**
  - left subarray, if $m < P$ — $\mathcal{E}_1 := \{m < P\}$
  - middle subarray, if $P < m < Q$ — $\mathcal{E}_2 := \{P < m < Q\}$
  - right subarray, if $Q < m$ — $\mathcal{E}_3 := \{Q < m\}$
  - or terminates if $m = P$ or $m = Q$

Notation

- Analyse **(random)** number of data **comparisons $C_n^{(m)}$** to select from random permutation

- **P** $<$ **Q**: (random) **ranks** of the two pivots

⇝ **subarray sizes** determined by P and Q:



$J_1 := P - 1$ $\quad J_2 := Q - P - 1$ $\quad J_3 := n - Q$

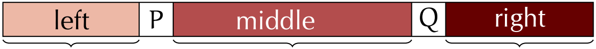
- selection continues recursively in — corresponding **events**
  - left subarray, if $m < P$ — $\mathcal{E}_1 := \{m < P\}$
  - middle subarray, if $P < m < Q$ — $\mathcal{E}_2 := \{P < m < Q\}$
  - right subarray, if $Q < m$ — $\mathcal{E}_3 := \{Q < m\}$
  - or terminates if $m = P$ or $m = Q$

Notation

- Analyse **(random)** number of data **comparisons** $C_n^{(m)}$ to select from random permutation

- **P** $<$ **Q**: (random) **ranks** of the two pivots

⇝ **subarray sizes** determined by P and Q:



$J_1 := P - 1$ $\quad$ $J_2 := Q - P - 1$ $\quad$ $J_3 := n - Q$

- selection continues recursively in
  - left subarray, if $m < P$
  - middle subarray, if $P < m < Q$
  - right subarray, if $Q < m$
  - or terminates if $m = P$ or $m = Q$

corresponding **events**

$\mathcal{E}_1 := \{m < P\}$

$\mathcal{E}_2 := \{P < m < Q\}$

$\mathcal{E}_3 := \{Q < m\}$

Quickselect Recurrence

- Yaroslavskiy's partitioning preserves randomness

⇝ can set up **recurrence** for the **distribution** of $C_n^{(m)}$:

$$C_n^{(m)} \overset{\mathcal{D}}{=} T_n + \begin{cases} C_{J_1}^{(m)} & \text{if } m \text{ lies in left subarray} \\ C_{J_2}^{(m-P)} & \text{if } m \text{ lies in middle subarray} \\ C_{J_3}^{(m-Q)} & \text{if } m \text{ lies in right subarray} \end{cases}$$

$$= T_n + \mathbb{1}_{\mathcal{E}_1} \cdot C_{J_1}^{(m)} + \mathbb{1}_{\mathcal{E}_2} \cdot C_{J_2}^{(m-P)} + \mathbb{1}_{\mathcal{E}_3} \cdot C_{J_3}^{(m-Q)}$$

- T_n is the (random) number of comparisons during **partitioning**

Getting Rid of m

- **Complication:** **second** parameter m in recurrence

$$C_n^{(m)} \overset{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1} \cdot C_{J_1}^{(m)} + \mathbb{1}_{\mathcal{E}_2} \cdot C_{J_2}^{(m-P)} + \mathbb{1}_{\mathcal{E}_3} \cdot C_{J_3}^{(m-Q)}$$

$\rightsquigarrow$ consider simpler quantities

1. **Random Rank:** $m = M_n \overset{\mathcal{D}}{=} \mathit{Uniform}\{1, \ldots, n\}$ ← *focus in talk*
   abbreviate $\bar{C}_n := C_n^{(M_n)}$

   $\rightsquigarrow$ $\bar{C}_n \overset{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1} \bar{C}_{J_1} + \mathbb{1}_{\mathcal{E}_2} \bar{C}_{J_2} + \mathbb{1}_{\mathcal{E}_3} \bar{C}_{J_3}$

2. **Extreme Rank:** $m = 1$
   abbreviate $\hat{C}_n := C_n^{(1)}$ $\rightsquigarrow$ $\hat{C}_n \overset{\mathcal{D}}{=} T_n + \hat{C}_{P-1}$

- explicit solution for **expectations** possible (generating functions)
- **But:** Requires tedious computations

*More elegant shortcut to **asymptotics** of $\mathbb{E}[\bar{C}_n]$ and $\mathbb{E}[\hat{C}_n]$?*

Getting Rid of m

- **Complication:** **second** parameter m in recurrence

$$C_n^{(m)} \stackrel{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1} \cdot C_{J_1}^{(m)} + \mathbb{1}_{\mathcal{E}_2} \cdot C_{J_2}^{(m-P)} + \mathbb{1}_{\mathcal{E}_3} \cdot C_{J_3}^{(m-Q)}$$

$\rightsquigarrow$ consider simpler quantities

1. **Random Rank:** $m = M_n \stackrel{\mathcal{D}}{=} \mathit{Uniform}\{1, \ldots, n\}$ $\leftarrow$ *focus in talk*

   abbreviate $\bar{C}_n := C_n^{(M_n)}$

   $$\rightsquigarrow \quad \boxed{\bar{C}_n \stackrel{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1} \bar{C}_{J_1} + \mathbb{1}_{\mathcal{E}_2} \bar{C}_{J_2} + \mathbb{1}_{\mathcal{E}_3} \bar{C}_{J_3}}$$

2. **Extreme Rank:** $m = 1$

   abbreviate $\hat{C}_n := C_n^{(1)}$ $\rightsquigarrow$ $\boxed{\hat{C}_n \stackrel{\mathcal{D}}{=} T_n + \hat{C}_{P-1}}$

- explicit solution for **expectations** possible (generating functions)
- **But:** Requires tedious computations

More elegant shortcut to ***asymptotics*** *of* $\mathbb{E}[\bar{C}_n]$ *and* $\mathbb{E}[\hat{C}_n]$*?*

Getting Rid of m

- **Complication:** **second** parameter m in recurrence

$$C_n^{(m)} \overset{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1} \cdot C_{J_1}^{(m)} + \mathbb{1}_{\mathcal{E}_2} \cdot C_{J_2}^{(m-P)} + \mathbb{1}_{\mathcal{E}_3} \cdot C_{J_3}^{(m-Q)}$$

$\rightsquigarrow$ consider simpler quantities

1. **Random Rank:** $m = M_n \overset{\mathcal{D}}{=} \mathit{Uniform}\{1,\ldots,n\}$ $\leftarrow$ *focus in talk*
   abbreviate $\bar{C}_n := C_n^{(M_n)}$

   $\rightsquigarrow$ $$\bar{C}_n \overset{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1}\bar{C}_{J_1} + \mathbb{1}_{\mathcal{E}_2}\bar{C}_{J_2} + \mathbb{1}_{\mathcal{E}_3}\bar{C}_{J_3}$$

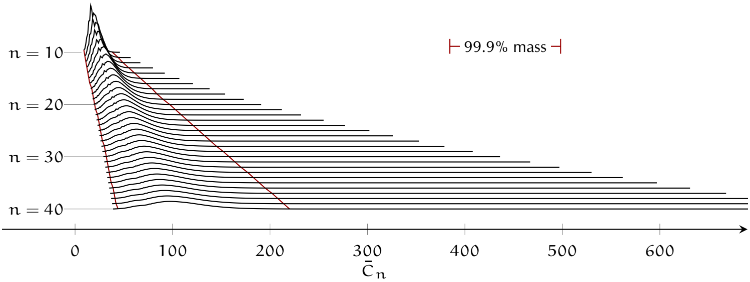
2. **Extreme Rank:** $m = 1$
   abbreviate $\hat{C}_n := C_n^{(1)}$ $\rightsquigarrow$ $$\hat{C}_n \overset{\mathcal{D}}{=} T_n + \hat{C}_{P-1}$$

- explicit solution for **expectations** possible (generating functions)
- **But:** Requires tedious computations

More elegant shortcut to ***asymptotics*** *of* $\mathbb{E}[\bar{C}_n]$ *and* $\mathbb{E}[\hat{C}_n]$*?*

Convergence

- **Idea:** Computation easy if we can **drop** index n
  i. e. when recurrence **"in equilibrium"**
- For that, we need stochastic **convergence** of $\bar{C}_n$



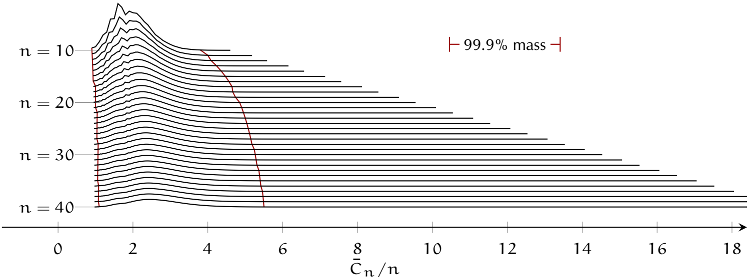
But:

- $\mathbb{E}[\bar{C}_n]$ grows with n
- high mass region moves with n

} $\bar{C}_n$ does **not** converge

Convergence

- **Idea:** Computation easy if we can **drop** index n
  i. e. when recurrence **"in equilibrium"**
- Try **normalized** variables $\bar{C}_n^* := \dfrac{\bar{C}_n}{n}$ instead:



- $\mathbb{E}[\bar{C}_n^*]$ seems constant
- range of high mass looks bounded

} $\bar{C}_n^*$ might converge

- Works because of property of Quickselect: $\mathbb{E}[\bar{C}_n] = \mathcal{O}\left(\sqrt{\mathrm{Var}(\bar{C}_n)}\right)$

The Contraction Method

- Rewrite recurrence: (Write as sum)

$$\bar{C}_n \overset{\mathcal{D}}{=} T_n + \mathbb{1}_{\mathcal{E}_1}\bar{C}_{J_1} + \mathbb{1}_{\mathcal{E}_2}\bar{C}_{J_2} + \mathbb{1}_{\mathcal{E}_3}\bar{C}_{J_3}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i}\frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}[|A_i^*|] < 1$

The Contraction Method

- Rewrite recurrence: (Divide by n)

$$\bar{C}_n \overset{\mathcal{D}}{=} T_n + \sum_{i=1}^{3} \mathbb{1}_{\mathcal{E}_i} \bar{C}_{J_i}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- Rewrite recurrence: (Rearrange)

$$\frac{\bar{C}_n}{n} \overset{\mathcal{D}}{=} \frac{T_n}{n} + \sum_{i=1}^{3} \frac{\mathbb{1}_{\mathcal{E}_i} \bar{C}_{J_i}}{n} \cdot \frac{J_i}{J_i}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{L_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- Rewrite recurrence: (use normalized variables)

$$\frac{\bar{C}_n}{n} \overset{\mathcal{D}}{=} \frac{T_n}{n} + \sum_{i=1}^{3} \mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \cdot \frac{\bar{C}_{J_i}}{J_i}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}[|A_i^*|] < 1$

The Contraction Method

- Recurrence for $\bar{C}^*_n$

$$\bar{C}^*_n \overset{\mathcal{D}}{=} T^*_n + \sum_{i=1}^{3} \mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \cdot \bar{C}^*_{J_i} \qquad \text{with } T^*_n := \frac{T_n}{n}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T^*_n \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A^*_i$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A^*_i|\big] < 1$

The Contraction Method

- Recurrence for $\bar{C}_n^*$

$$\bar{C}_n^* \overset{\mathcal{D}}{=} \underbrace{T_n^*}_{\downarrow \; T^*} + \sum_{i=1}^{3} \underbrace{\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n}}_{\downarrow \; A_i^*} \cdot \bar{C}_{J_i}^*$$

- **Contraction Method:**
  Convergence of coefficients *+ contraction condition implies convergence to fixpoint solution.*

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- Recurrence for $\bar{C}^*_n$

$$\bar{C}^*_n \overset{\mathcal{D}}{=} \underbrace{T^*_n}_{\downarrow\, T^*} + \sum_{i=1}^{3} \underbrace{\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n}}_{\downarrow\, A^*_i} \cdot \bar{C}^*_{J_i}$$

$$\sum_{i=1}^{3} \mathbb{E}\big[|A^*_i|\big] < 1$$

- **Contraction Method:**
  *Convergence of coefficients + **contraction condition** implies convergence to fixpoint solution.*

⇝ have to show: (all convergences in L_1-norm)

1. $T^*_n \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A^*_i$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A^*_i|\big] < 1$

The Contraction Method

- Recurrence for $\bar{C}_n^*$

$$\begin{array}{ccccccc} \bar{C}_n^* & \overset{\mathcal{D}}{=} & T_n^* + \sum_{i=1}^{3} & \underbrace{\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n}} & \cdot & \bar{C}_{J_i}^* \\ \downarrow & & \downarrow & \downarrow & & \downarrow \\ \bar{C}^* & \overset{\mathcal{D}}{=} & T^* + \sum_{i=1}^{3} & A_i^* & \cdot & \bar{C}^* \end{array}$$

$$\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition
  implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- **"Equilibrium" equation for $\bar{C}^*$**

$$\bar{C}^* \overset{\mathcal{D}}{=} T^* + \sum_{i=1}^{3} A_i^* \cdot \bar{C}^*$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- **"Equilibrium" equation for $\bar{C}^*$**

(Take **expectation** on both sides)

$$\bar{C}^* \overset{\mathcal{D}}{=} T^* + \sum_{i=1}^{3} A_i^* \cdot \bar{C}^*$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- **"Equilibrium" equation for $\bar{C}^*$**

(and **solve** for $\mathbb{E}\,\bar{C}^*$)

$$\mathbb{E}\,\bar{C}^* = \mathbb{E}\,T^* + \sum_{i=1}^{3} \mathbb{E}\,A_i^* \cdot \mathbb{E}\,\bar{C}^*$$

- **Contraction Method:**
Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

$\rightsquigarrow$ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- **"Equilibrium" equation for $\bar{C}^*$**

  (and **solve** for $\mathbb{E}\,\bar{C}^*$)

$$\mathbb{E}\,\bar{C}^* = \frac{\mathbb{E}\,T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\,A_i^*}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

$\rightsquigarrow$ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

The Contraction Method

- **"Equilibrium" equation for $\bar{C}^*$**

  (and **solve** for $\mathbb{E}\,\bar{C}^*$)

$$\mathbb{E}\,\bar{C}^* \;=\; \frac{\mathbb{E}\,T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\,A_i^*}$$

- **Contraction Method:**
  Convergence of coefficients + contraction condition implies convergence to fixpoint solution.

⇝ have to show: (all convergences in L_1-norm)

1. $T_n^* \to T^*$
2. $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$ … *upcoming*
3. $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

Probabilistic Model

Need "convenient" probabilistic model:
same probability space *for finite n and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d. *Uniform*(0, 1)**

first two = pivots values

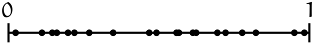
- **Our equivalent model:**

  1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
  3. Continue recursively

Probabilistic Model

Need "convenient" probabilistic model:
same probability space *for finite n and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d. *Uniform*(0, 1)**



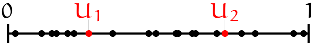
first two = pivots values

- **Our equivalent model:**
  1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
  3. Continue recursively

Probabilistic Model

Need "convenient" probabilistic model:
same probability space *for finite n and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$



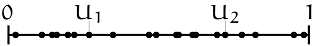
first two = pivots values

- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $S \overset{\mathcal{D}}{=} \textit{Dirichlet}(1, 1, 1)$
2. Draw **subarray sizes** I_1, I_2, I_3 with $I \overset{\mathcal{D}}{=} \textit{Multinomial}(n-2; S_1, S_2, S_3)$
3. Continue recursively

Need "convenient" probabilistic model:
same probability space *for finite n and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$



first two $=$ pivots values

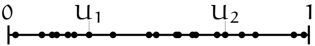
- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=}$ *Dirichlet*$(1, 1, 1)$
2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=}$ *Multinomial*$(n - 2; S_1, S_2, S_3)$
3. Continue recursively

Probabilistic Model

Need "convenient" probabilistic model:
same probability space *for finite* n *and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$



first two $=$ pivots values

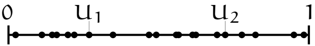
- **Our equivalent model:**
  1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n - 2; S_1, S_2, S_3)$
  3. Continue recursively

- induces **pivot values:** $U_1 = S_1$, $U_2 = S_1 + S_2$

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

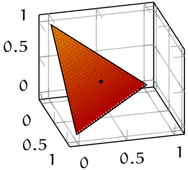
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$



first two $=$ pivots values

- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=}$ *Dirichlet*$(1, 1, 1)$
2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=}$ *Multinomial*$(n-2; S_1, S_2, S_3)$
3. Continue recursively

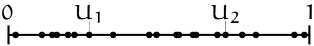


- induces **pivot values:** $U_1 = S_1$, $U_2 = S_1 + S_2$

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

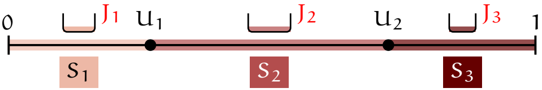
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$

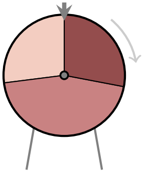


first two = pivots values

- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n - 2; S_1, S_2, S_3)$
3. Continue recursively



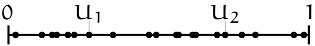


- $S_1 = \Pr(\text{small})$, $S_2 = \Pr(\text{medium})$, $S_3 = \Pr(\text{large})$
- $n - 2$ elements to draw **conditional on S**

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

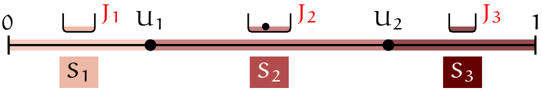
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d. *Uniform***$(0,1)$

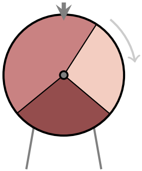


first two $=$ pivots values

- **Our equivalent model:**

  1. Draw **spacings** S_1, S_2, S_3 on $(0,1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1,1,1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
  3. Continue recursively



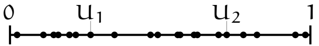


- $S_1 = \Pr(\text{small})$, $S_2 = \Pr(\text{medium})$, $S_3 = \Pr(\text{large})$
- $n-2$ elements to draw **conditional on S**

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

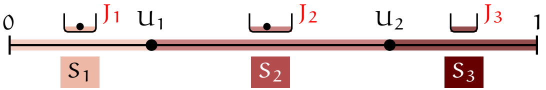
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$

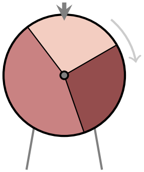


first two $=$ pivots values

- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \textit{Dirichlet}(1, 1, 1)$
2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \textit{Multinomial}(n-2; S_1, S_2, S_3)$
3. Continue recursively



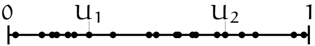


- $S_1 = \Pr(\text{small})$, $S_2 = \Pr(\text{medium})$, $S_3 = \Pr(\text{large})$
- $n-2$ elements to draw **conditional on S**

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

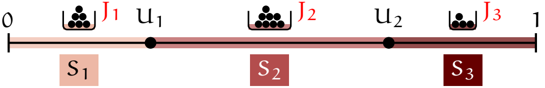
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d. *Uniform*$(0, 1)$**

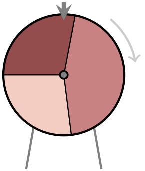


first two $=$ pivots values

- **Our equivalent model:**

1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
3. Continue recursively



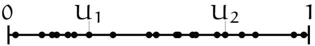


- $S_1 = \Pr(\text{small})$, $S_2 = \Pr(\text{medium})$, $S_3 = \Pr(\text{large})$
- $n - 2$ elements to draw **conditional on S**

Probabilistic Model

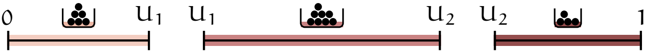
Need "convenient" probabilistic model:
same probability space *for finite n and limit case*

- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d.** ***Uniform***$(0, 1)$



first two = pivots values

- **Our equivalent model:**
  1. Draw **spacings** S_1, S_2, S_3 on $(0, 1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1, 1, 1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
  3. Continue recursively

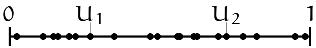


- Continue of sub-interval with smaller n

Probabilistic Model

Need "convenient" probabilistic model:
***same probability space** for finite n and limit case*

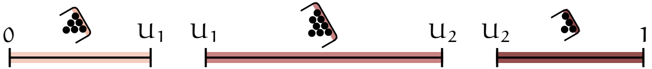
- **Natural model:** input elements $U_1, \ldots, U_n$ **i. i. d. *Uniform*$(0,1)$**



first two = pivots values

- **Our equivalent model:**

  1. Draw **spacings** S_1, S_2, S_3 on $(0,1)$ with $\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(1,1,1)$
  2. Draw **subarray sizes** J_1, J_2, J_3 with $\mathbf{J} \overset{\mathcal{D}}{=} \mathit{Multinomial}(n-2; S_1, S_2, S_3)$
  3. Continue recursively



- Continue of sub-interval with smaller n

Distribution of T_n — Repetition

❶ $T_n^* \to T^*$

T_n = number of comparisons in **first partitioning step**

$\sim$	n	one for every element
$+$	J_2	second for **all** medium elements
$+$	$l@\mathcal{K}$	second for large elements at C'_k
$+$	$s@\mathcal{G}$	second for small elements at C'_g

- Markus Nebel just showed:

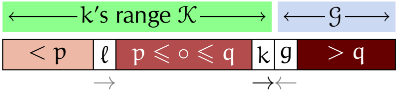
$$l@\mathcal{K} \stackrel{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\, J_3,\, J_1+J_2)$$
$$s@\mathcal{G} \stackrel{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\, J_1,\, J_3)$$

} **conditional** on **J**

Distribution of T_n — Repetition

❶ $T_n^* \to T^*$

Invariant:



$$T_n \;=\; \text{number of comparisons in } \textbf{first partitioning step}$$

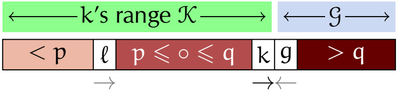
$\sim$	n	one for every element
	$+\; J_2$	second for **all** medium elements
	$+\; l@\mathcal{K}$	second for large elements at c'_k
	$+\; s@\mathcal{G}$	second for small elements at c'_g

- Markus Nebel just showed:

$$\left.\begin{aligned} l@\mathcal{K} &\overset{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\, J_3,\, J_1+J_2) \\ s@\mathcal{G} &\overset{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\, J_1,\, J_3) \end{aligned}\right\}\ \text{conditional on } \mathbf{J}$$

Distribution of T_n — Repetition

❶ $T_n^* \to T^*$

Invariant: 

$$
\begin{aligned}
T_n &= \text{number of comparisons in } \textbf{first partitioning step} \\
&\sim \quad n \qquad && \text{one for every element} \\
&\quad + J_2 && \text{second for } \textbf{all} \text{ medium elements} \\
&\quad + l@\mathcal{K} && \text{second for large elements at } c'_k \\
&\quad + s@\mathcal{G} && \text{second for small elements at } c'_g
\end{aligned}
$$

- Markus Nebel just showed:

$$
\left.
\begin{aligned}
l@\mathcal{K} &\overset{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\ J_3,\ J_1 + J_2) \\
s@\mathcal{G} &\overset{\mathcal{D}}{=} \mathit{Hypergeometric}(n-2;\ J_1,\ J_3)
\end{aligned}
\right\} \text{ conditional on } \mathbf{J}
$$

Convergence of T_n^*

❶ $T_n^* \to T^*$

$$T_n^* = \frac{T_n}{n} = 1 + \frac{J_2}{n} + \frac{l@\mathcal{K}}{n} + \frac{s@\mathcal{G}}{n}$$

- $\frac{J_2}{n} \to S_2$ by **law of large numbers**, since $\mathbb{E}[J_i \mid \mathbf{S}] = n \cdot S_i$
- Multinomial and Hypergeometric **concentrated** around mean

$$\rightsquigarrow \frac{l@\mathcal{K}}{n} \to \mathbb{E}\left[\frac{l@\mathcal{K}}{n} \,\middle|\, \mathbf{S}\right] \quad \text{by **Chernoff bound** arguments}$$

$$= \mathbb{E}\left[\frac{\mathbb{E}[l@\mathcal{K} \mid \mathbf{J}]}{n} \,\middle|\, \mathbf{S}\right] = \mathbb{E}\left[\frac{J_3(J_1+J_2)}{n(n-2)} \,\middle|\, \mathbf{S}\right]$$

$$\sim S_3 \cdot (S_1 + S_2)$$

- $\frac{s@\mathcal{G}}{n} \to S_1 \cdot S_3$ similar

$$\rightsquigarrow T_n^* \to T^* = 1 + S_2 + S_3(2S_1 + S_2)$$

Convergence of T_n^*

❶ $T_n^* \to T^*$

$$T_n^* = \frac{T_n}{n} = 1 + \frac{J_2}{n} + \frac{l@\mathcal{K}}{n} + \frac{s@\mathcal{G}}{n}$$

- $\frac{J_2}{n} \to S_2$ by **law of large numbers**, since $\mathbb{E}[J_i \mid \mathbf{S}] = n \cdot S_i$
- Multinomial and Hypergeometric **concentrated** around mean

$\rightsquigarrow$ $\frac{l@\mathcal{K}}{n} \to \mathbb{E}\left[\frac{l@\mathcal{K}}{n} \,\middle|\, \mathbf{S}\right]$ by **Chernoff bound** arguments

$$= \mathbb{E}\left[\frac{\mathbb{E}[l@\mathcal{K} \mid \mathbf{J}]}{n} \,\middle|\, \mathbf{S}\right] = \mathbb{E}\left[\frac{J_3(J_1 + J_2)}{n(n-2)} \,\middle|\, \mathbf{S}\right]$$

$$\sim S_3 \cdot (S_1 + S_2)$$

- $\frac{s@\mathcal{G}}{n} \to S_1 \cdot S_3$ similar

$\rightsquigarrow$ $T_n^* \to T^* = 1 + S_2 + S_3(2S_1 + S_2)$

Convergence of T_n^*

❶ $T_n^* \to T^*$

$$T_n^* = \frac{T_n}{n} = 1 + \frac{J_2}{n} + \frac{l@\mathcal{K}}{n} + \frac{s@\mathcal{G}}{n}$$

- $\frac{J_2}{n} \to S_2$ by **law of large numbers**, since $\mathbb{E}[J_i \mid \mathbf{S}] = n \cdot S_i$
- Multinomial and Hypergeometric **concentrated** around mean

$$\rightsquigarrow \frac{l@\mathcal{K}}{n} \to \mathbb{E}\left[\frac{l@\mathcal{K}}{n} \,\middle|\, \mathbf{S}\right] \text{ by \textbf{Chernoff bound} arguments}$$

$$= \mathbb{E}\left[\frac{\mathbb{E}\left[l@\mathcal{K} \mid \mathbf{J}\right]}{n} \,\middle|\, \mathbf{S}\right] = \mathbb{E}\left[\frac{J_3(J_1+J_2)}{n(n-2)} \,\middle|\, \mathbf{S}\right]$$

$$\sim S_3 \cdot (S_1 + S_2)$$

- $\frac{s@\mathcal{G}}{n} \to S_1 \cdot S_3$ similar

$$\rightsquigarrow T_n^* \to T^* = 1 + S_2 + S_3(2S_1 + S_2)$$

Convergence of Coefficients & Contraction Condition

2 $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$

- standard computation:
  $$\mathbb{1}_{\mathcal{E}_1} \frac{J_1}{n} \;\to\; \mathbb{1}_{\mathcal{F}_1} \cdot S_1 \quad \text{with } \mathcal{F}_1 = \{V < S_1\}$$
- $i = 2, 3$ similar

3 $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

- $\mathbb{E}\big[\mathbb{1}_{\mathcal{F}_1} S_1\big] = \frac{1}{6}$

⇝ by symmetry: $\displaystyle\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] = \frac{1}{2} < 1$

⇝ ***We (finally) proved convergence*** $\bar{C}_n^* \to \bar{C}^*$.

Convergence of Coefficients & Contraction Condition

2 $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$

- standard computation:
  $$\mathbb{1}_{\mathcal{E}_1} \frac{J_1}{n} \;\to\; \mathbb{1}_{\mathcal{F}_1} \cdot S_1 \quad \text{with } \mathcal{F}_1 = \{V < S_1\}$$
- $i = 2, 3$ similar

3 $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

- $\mathbb{E}\big[\mathbb{1}_{\mathcal{F}_1} S_1\big] = \frac{1}{6}$

⇝ by symmetry: $\displaystyle\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] = \frac{1}{2} < 1$

⇝ *We (finally) proved convergence* $\bar{C}_n^* \to \bar{C}^*$.

Convergence of Coefficients & Contraction Condition

2 $\mathbb{1}_{\mathcal{E}_i} \frac{J_i}{n} \to A_i^*$

- standard computation:
  $$\mathbb{1}_{\mathcal{E}_1} \frac{J_1}{n} \;\to\; \mathbb{1}_{\mathcal{F}_1} \cdot S_1 \quad \text{with } \mathcal{F}_1 = \{V < S_1\}$$
- $i = 2, 3$ similar

3 $\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] < 1$

- $\mathbb{E}\big[\mathbb{1}_{\mathcal{F}_1} S_1\big] = \frac{1}{6}$

$\leadsto$ by symmetry: $\displaystyle\sum_{i=1}^{3} \mathbb{E}\big[|A_i^*|\big] = \frac{1}{2} < 1$

$\leadsto$ ***We (finally) proved convergence*** $\bar{C}_n^* \to \bar{C}^*$.

Results — Expectations

Computing expectations and plugging in . . .

- $\mathbb{E}\,T^* = \frac{19}{12}$
- $\sum_{i=1}^{3} \mathbb{E}\,A_i^* = \frac{1}{2}$

$\Bigg\} \rightsquigarrow \mathbb{E}\,\bar{C}^* = \frac{19}{6} = 3.1\overline{6}$

Recall:

$$\mathbb{E}\,\bar{C}^* = \frac{\mathbb{E}\,T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\,A_i^*}$$

- L_1 convergence $\rightsquigarrow$ $\mathbb{E}\,\bar{C}_n \sim 3.1\overline{6}n$
- similarly: $\mathbb{E}\,\hat{C}_n \sim 2.375n$

Results — Expectations

Computing expectations and plugging in . . .

- $\mathbb{E}\, T^* = \frac{19}{12}$
- $\sum_{i=1}^{3} \mathbb{E}\, A_i^* = \frac{1}{2}$

$\Bigg\} \rightsquigarrow \mathbb{E}\, \bar{C}^* = \frac{19}{6} = 3.1\overline{6}$

Recall:

$$\mathbb{E}\, \bar{C}^* = \frac{\mathbb{E}\, T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\, A_i^*}$$

- L_1 convergence $\rightsquigarrow$ $\mathbb{E}\, \bar{C}_n \sim 3.1\overline{6}n$
- similarly: $\mathbb{E}\, \hat{C}_n \sim 2.375n$

Computing expectations and plugging in ...

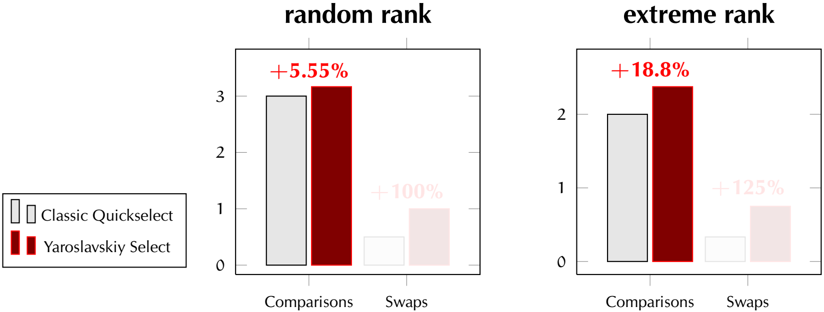
- $\mathbb{E}\,T^* = \frac{19}{12}$
- $\sum_{i=1}^{3} \mathbb{E}\,A_i^* = \frac{1}{2}$

$\Big\} \rightsquigarrow \mathbb{E}\,\bar{C}^* = \frac{19}{6} = 3.1\overline{6}$

Recall:

$$\mathbb{E}\,\bar{C}^* = \frac{\mathbb{E}\,T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\,A_i^*}$$

- L_1 convergence $\rightsquigarrow$ $\mathbb{E}\,\bar{C}_n \sim 3.1\overline{6}n$
- similarly: $\mathbb{E}\,\hat{C}_n \sim 2.375n$



Computing expectations and plugging in ...

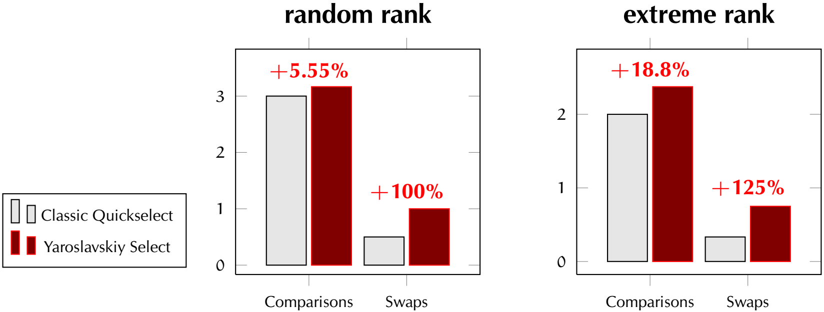
- $\mathbb{E}\, T^* = \frac{19}{12}$
- $\sum_{i=1}^{3} \mathbb{E}\, A_i^* = \frac{1}{2}$

$\Bigg\} \rightsquigarrow \mathbb{E}\, \bar{C}^* = \frac{19}{6} = 3.1\overline{6}$

Recall:

$$\mathbb{E}\, \bar{C}^* = \frac{\mathbb{E}\, T^*}{1 - \sum_{i=1}^{3} \mathbb{E}\, A_i^*}$$

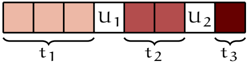
- L_1 convergence $\rightsquigarrow$ $\mathbb{E}\, \bar{C}_n \sim 3.1\overline{6}n$
- similarly: $\mathbb{E}\, \hat{C}_n \sim 2.375n$



Pivot Sampling

- successful optimization of Quicksort/-select: **median of three**
- **Here:** generalized pivot selection scheme
  1. Choose random sample of k elements
  2. **Sort** the sample
  3. Select pivots s. t. in sorted sample
     t_1 smaller,
     t_2 between and
     t_3 larger than pivots

Example with
$k = 8$ and $\mathbf{t} = (3, 2, 1)$:



Our stochastic model nicely **generalizes** to pivot sampling:

- **only** the distribution of spacings $\mathbf{S}$ changes!

$$\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(t_1 + 1, t_2 + 1, t_3 + 1) \qquad \textit{density: } f(s) = \frac{s_1^{t_1} s_2^{t_2} s_3^{t_3}}{B}$$

- contraction conditions remain valid

Pivot Sampling

- successful optimization of Quicksort/-select: **median of three**
- **Here:** generalized pivot selection scheme
  1. Choose random sample of k elements
  2. **Sort** the sample
  3. Select pivots s. t. in sorted sample

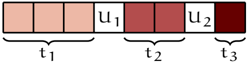
     t_1 smaller,

     t_2 between and

     t_3 larger than pivots

Example with

$k = 8$ and $\mathbf{t} = (3, 2, 1)$:



Our stochastic model nicely **generalizes** to pivot sampling:

- **only** the **distribution of spacings** $\mathbf{S}$ changes!

$$\mathbf{S} \overset{\mathcal{D}}{=} \mathit{Dirichlet}(t_1 + 1,\ t_2 + 1,\ t_3 + 1) \qquad \textit{density: } f(s) = \frac{s_1^{t_1} s_2^{t_2} s_3^{t_3}}{B}$$

- contraction conditions remain valid

Results — Pivot Sampling

- **General result:** $\bar{C}_n \sim \dfrac{1 + \frac{t_2+1}{k+1} + \frac{(2(t_1+1)+(t_2+1))(t_3+1)}{(k+1)^{\overline{2}}}}{1 - \sum_{i=1}^{3} \frac{(t_i+1)^{\overline{2}}}{(k+1)^{\overline{2}}}} \cdot n$

  inconvenient for interpretation . . .

- **Limiting case for large sample:**
  - $\frac{t_i}{k} \to \alpha_i$ as $k \to \infty$

  $\triangleq$ take exact quantiles as pivots

$$\bar{C}_n \sim \frac{1 + \alpha_2 + (2\alpha_1 + \alpha_2)\alpha_3}{1 - \sum_{i=1}^{3} \alpha_i^2} \cdot n$$

$\rightsquigarrow$ optimal choice is not symmetric

- classic Quickselect with median of $2t+1$:

  $2.5n$ for $t = 1$ and $2.3n$ for $t = 2$

Results — Pivot Sampling

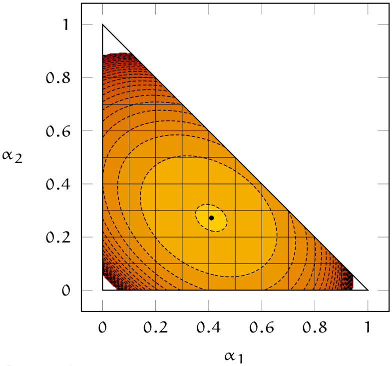
- **General result:** $\bar{C}_n \sim \dfrac{1 + \frac{t_2+1}{k+1} + \frac{(2(t_1+1)+(t_2+1))(t_3+1)}{(k+1)^{\overline{2}}}}{1 - \sum_{i=1}^{3} \frac{(t_i+1)^{\overline{2}}}{(k+1)^{\overline{2}}}} \cdot n$

  inconvenient for interpretation ...

- **Limiting case for large sample:**
  - $\frac{t_i}{k} \to \alpha_i$ as $k \to \infty$

  $\triangleq$ take exact quantiles as pivots

$$\bar{C}_n \sim \frac{1 + \alpha_2 + (2\alpha_1 + \alpha_2)\alpha_3}{1 - \sum_{i=1}^{3} \alpha_i^2} \cdot n$$



$\rightsquigarrow$ optimal choice is **not symmetric**

- **classic Quickselect** with **median** of $2t+1$:
  $2.5n$ for $t = 1$ and $2.\overline{3}n$ for $t = 2$

Results — Pivot Sampling

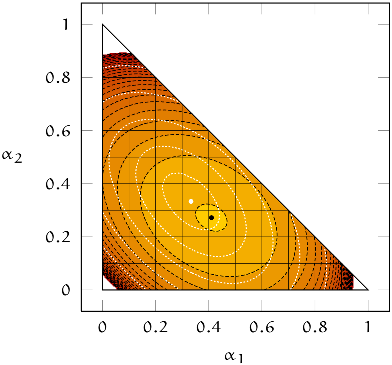
- **General result:** $\bar{C}_n \sim \dfrac{1 + \frac{t_2+1}{k+1} + \frac{(2(t_1+1)+(t_2+1))(t_3+1)}{(k+1)^{\overline{2}}}}{1 - \sum_{i=1}^{3} \frac{(t_i+1)^{\overline{2}}}{(k+1)^{\overline{2}}}} \cdot n$

  inconvenient for interpretation . . .

- **Limiting case for large sample:**
  - $\dfrac{t_i}{k} \to \alpha_i$ as $k \to \infty$

  $\triangleq$ take exact quantiles as pivots

$$\bar{C}_n \sim \frac{1 + \alpha_2 + (2\alpha_1 + \alpha_2)\alpha_3}{1 - \sum_{i=1}^{3} \alpha_i^2} \cdot n$$



$\rightsquigarrow$ optimal choice is **not symmetric**

- **classic Quickselect** with **median** of $2t + 1$:
  $2.5n$ for $t = 1$ and $2.\overline{3}n$ for $t = 2$

Results — Pivot Sampling

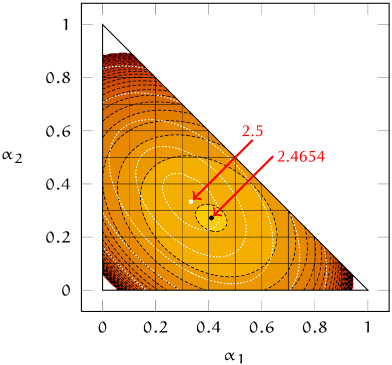
- **General result:** $\bar{C}_n \sim \dfrac{1 + \frac{t_2+1}{k+1} + \frac{(2(t_1+1)+(t_2+1))(t_3+1)}{(k+1)^{\overline{2}}}}{1 - \sum_{i=1}^{3} \frac{(t_i+1)^{\overline{2}}}{(k+1)^{\overline{2}}}} \cdot n$

  inconvenient for interpretation . . .

- **Limiting case for large sample:**
  - $\dfrac{t_i}{k} \to \alpha_i$ as $k \to \infty$

  $\triangleq$ take exact quantiles as pivots

$$\bar{C}_n \sim \frac{1 + \alpha_2 + (2\alpha_1 + \alpha_2)\alpha_3}{1 - \sum_{i=1}^{3} \alpha_i^2} \cdot n$$



$\rightsquigarrow$ optimal choice is **not symmetric**

- **classic Quickselect** with **median** of $2t + 1$:
  $2.5n$ for $t = 1$ and $2.\overline{3}n$ for $t = 2$

Conclusion

Practitioners' Lesson:

- Yaroslavskiy's dual-pivoting **not** helpful for selection

Analyzers' Lesson:

- **expected** costs of Quickselect derivable by **contraction method**
- pivot sampling included naturally

References & Further Reading

M. Blum, R. W. Floyd, V. Pratt, R. L. Rivest and R. E. Tarjan.
Time bounds for selection.
J. of Comp. and System Sc., 7(4):448–461, 1973.

C. A. R. Hoare.
Algorithm 65: Find.
Communications of the ACM, 4(7):321–322, July 1961.

H.-K. Hwang and T.-H. Tsai.
Quickselect and Dickman function.
Combinatorics, Probability and Computing, 11, 2000.

Hosam M. Mahmoud.
Sorting: A distribution theory.
John Wiley & Sons, Hoboken, NJ, USA, 2000.

H. M. Mahmoud, R. Modarres and R. T. Smythe.
Analysis of quickselect : an algorithm for order statistics.
Informatique théorique et appl., 29(4):255–276, 1995.

C. Martínez, D. Panario and A. Viola.
Adaptive sampling strategies for quickselects.
ACM Transactions on Algorithms, 6(3):1–45, June 2010.

C. Martínez and S. Roura.
Optimal Sampling Strategies in Quicksort and Quickselect.
SIAM Journal on Computing, 31(3):683, 2001.

R. Neininger.
On a multivariate contraction method for random recursive structures with applications to Quicksort.
Random Structures & Alg., 19(3-4):498–524, 2001.

R. Neininger and L. Rüschendorf.
A General Limit Theorem for Recursive Algorithms and Combinatorial Structures.
The Annals of Applied Probability, 14(1):378–418, 2004.

A. Panholzer and H. Prodinger.
A generating functions approach for the analysis of grand averages for multiple QUICKSELECT.
Random Structures & Alg., 13(3-4):189–209, 1998.

U. Rösler and L. Rüschendorf.
The contraction method for recursive algorithms.
Algorithmica, 29(1):3–33, 2001.

S. Wild.
Java 7's Dual Pivot Quicksort.
Master thesis, University of Kaiserslautern, 2012.

S. Wild, M. E. Nebel and H. Mahmoud.
Analysis of Quickselect under Yaroslavskiy's Dual-Pivoting Algorithm.
***Algorithmica*, in preparation.**

S. Wild, M. E. Nebel and R. Neininger.
Average Case and Distributional Analysis of Java 7's Dual Pivot Quicksort.
ACM Transactions on Algorithms, submitted.

V. Yaroslavskiy.
Dual-Pivot Quicksort.
2009.

Results — Variances

- more diligence in analysis: convergences hold in L_2

$\rightsquigarrow$ $\mathrm{Var}(\bar{C}^*_n) \to \mathrm{Var}(\bar{C}^*)$ and $\mathbf{Var}(\bar{C}_n) \sim \mathbf{Var}(\bar{C}^*)n^2$

	random rank	**extreme rank**	
		min	max
classic Quickselect	1	0.5	
Yaroslavskiy Select	0.694	0.263	0.358

- variances **lower** than for classic Quickselect
- **asymmetry** for extreme ranks:
  - T^* **not** symmetric in $\mathbf{S}$

  $\rightsquigarrow$ Cost of selecting minimum and maximum have **different distribution** (but same mean)

Results — Variances

- more diligence in analysis: convergences hold in L_2

$\rightsquigarrow$ $\mathrm{Var}(\bar{C}_n^*) \to \mathrm{Var}(\bar{C}^*)$ and $\mathbf{Var}(\bar{C}_n) \sim \mathbf{Var}(\bar{C}^*) n^2$

	random rank	**extreme rank**	
		min	max
classic Quickselect	1	0.5	
Yaroslavskiy Select	0.694	0.263	0.358

- variances **lower** than for classic Quickselect
- **asymmetry** for extreme ranks:
  - T^* **not** symmetric in **S**

  $\rightsquigarrow$ Cost of selecting minimum and maximum have **different distribution** (but same mean)