

Algorithms for Stream Processing

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1
1	$\frac{1}{2}$	2

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1
1	$\frac{1}{2}$	2
2	$\frac{1}{4}$	4

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1
1	$\frac{1}{2}$	2
2	$\frac{1}{4}$	4
3	$\frac{1}{8}$	8

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1
1	$\frac{1}{2}$	2
2	$\frac{1}{4}$	4
3	$\frac{1}{8}$	8
4	$\frac{1}{16}$	16

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

n	Chance to increment	Estimation
0	1	1
1	$\frac{1}{2}$	2
2	$\frac{1}{4}$	4
3	$\frac{1}{8}$	8
4	$\frac{1}{16}$	16
...
n	$\frac{1}{2^n}$	2^n

Morris Counter (1977)



Robert Morris (July 25, 1932 – June 26, 2011)

```
def probably_increment(x: Element): Unit = {  
    // need edge case because generating  
    // random of 0 is an error  
    if (counter == 0) {  
        counter += 1  
        return  
    }  
  
    //increment with probability 1 / 2^counter  
    val r = Random.nextInt(pow(2, counter))  
    if (r == 0) {  
        counter += 1  
    }  
}
```


Hashing

$hash(x) \sim \{0,1\}^4$

$hash("foobar") = 1011$ // von zufälliger Folge nicht zu unterscheiden

Wie oft muss man im Durchschnitt würfeln, um eine 6 zu bekommen? // W6

6 mal

Wie viele unterschiedliche Werte muss man im Durchschnitt hashen um alle Kombinationen zu erhalten? // 4-bit Hash

$2^4 = 16$

Wie viele unterschiedliche Werte muss man im Durchschnitt hashen um einen bestimmten Hash zu erhalten? // z.B. 1010

$2^4 = 16$

HyperLogLog (2007)



Philippe Flajolet
Olivier Gandouet

Éric Fusy,
Frédéric Meunier

Wie viele unterschiedliche Werte muss man im Durchschnitt Hashen um drei führende Nullen zu erhalten?

00010101011101101

$$2^3 = 8$$

Wie viele führenden Nullen hat man maximal, wenn man 1024 unterschiedliche Werte hasht?

000000000000101001

$$\log_2(1024) = 10$$

Und für n Bit?

$$\log_2(n)$$

Varianz

Beispiel: MorrisCounter

$$n = 63 \rightarrow 2^{63} \approx 9.000.000.000.000.000.000$$

→ Man braucht $\sim 2^{63}$ weitere Versuche um n zu erhöhen

→ Das ist nicht mal mehr auf die Milliarde genau

→ Immer ungenauer für größere Zahlen

Analoges Problem bei HyperLogLog

1
2
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
131.072
262.144
524.288

Lösung für das Varianzproblem

Mehrere Counter parallel berechnen und einen Mittelwert bilden.

Beispiel:

	Wahrer Wert	Einfach	Artith. Mittel	Harm. Mittel
Counter	1.117	2.047	1.379	1.069
Distinct Elements	321	512	939	293

1
2
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
131.072
262.144
524.288