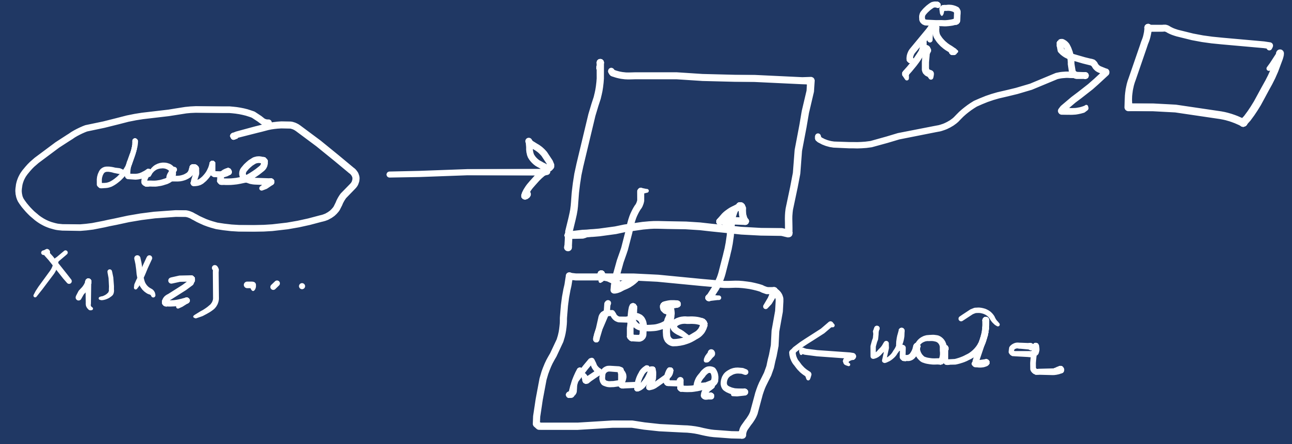
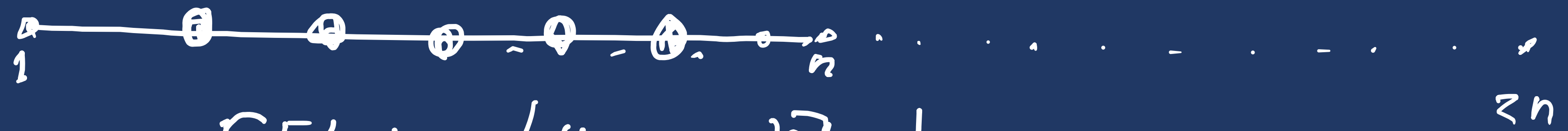
STREAMING

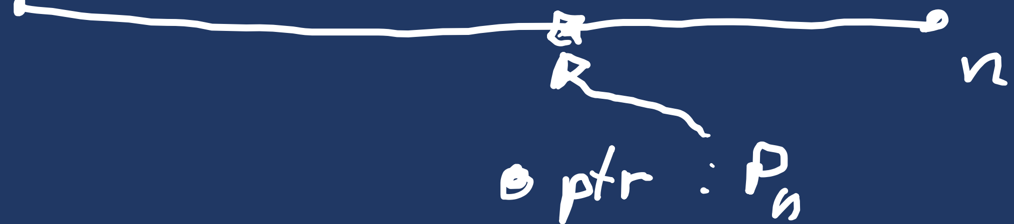


Generowanie próbki losowej.



CEL: jedn. rozkład

CEL: próbka 1-elementowa



$(\forall n)$ P_n jest jedn. rozł. na $\{1, .., n\}$

$$(\forall n)(\forall i \in \{1..n\})\left(\Pr[P_n = i] = \frac{1}{n}\right)$$

Algorytm R (Vitter)

```
onInit () {
    n = ∅ ; // licznik
    p = nil ; // ~~[illegible]~~ wskaźnik
  data = nil ; // kopia danych
}

OnRead (x) { n++ ;
    if (random(0,1) < 1/n) {
        p = n ;
        data = x
    }
}

on Get () {
  return ((p, data))
}
```

generator.seed(10 [illegible])

p_n = wartość wskaźnika po przeczytaniu n-tego elem.,

$p_n \in \{1, \dots, n\}$

ZAŁ. $P[p_n = i] = \frac{1}{n} \quad \forall i \in \{1, \dots, n\}$

$$P[p_{n+1} = i] = P[p_{n+1} = i \mid Z] \cdot P[Z] + P[p_{n+1} = i \mid \neg Z] \cdot P[\neg Z]$$

Z = "zaszła zmiana w momencie n+1"

1. $i \in \{1, \dots, n\}$

$$P[p_{n+1} = i] = 0 \cdot \frac{1}{n+1} + P[p_n = i] \cdot \left(1 - \frac{1}{n+1}\right) =$$
$$= \frac{1}{n}\left(1 - \frac{1}{n+1}\right) = \frac{1}{n} \frac{n}{n+1} = \frac{1}{n+1}$$

2. $i = n+1$

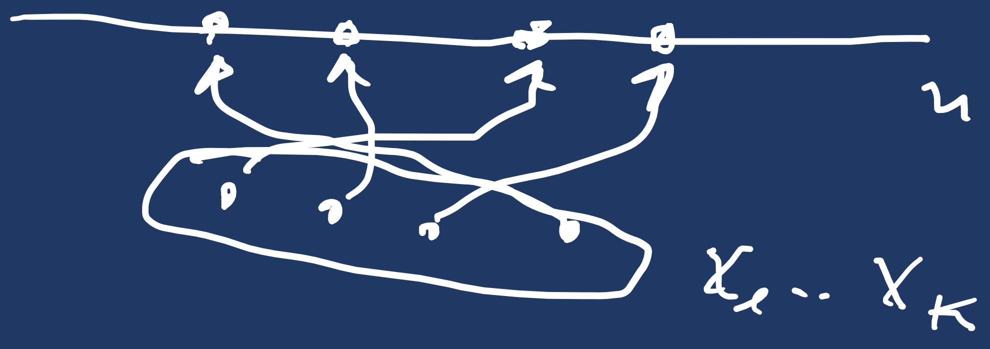
$$P[p_{n+1} = i] = 1 \cdot \frac{1}{n+1} + 0 = \frac{1}{n+1}$$

$\blacksquare$

Co zrobić aby otrzym. próbkę rozmiaru k ($k \ll n$)

① Generujemy niezależnie k liczników:

$$X_1, \dots, X_k$$



$$P\left[\bigwedge_{1 \le i < j \le k} \left(X_i^{(n)} \ne X_j^{(n)}\right)\right] =$$

$$= 1 - P\left[\bigvee_{i<j} X_i^{(n)} = X_j^{(n)}\right] \ge 1 - \sum_{1 \le i < j \le k} P\left[X_i^{(n)} = X_j^{(n)}\right] =$$

B.P.

$$= 1 - \frac{k(k-1)}{2}\frac{1}{n} \simeq 1 - \frac{k^2}{2n}.$$

wn. $n \gg k^2/2$

$\Downarrow$

próbka jest różna k w.h.p

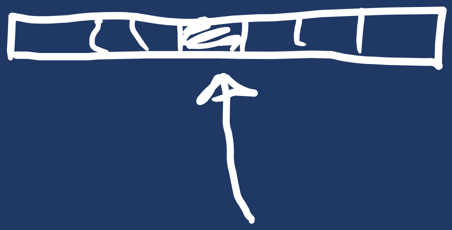
2. Chcemy wygen $[X_1, .., X_k]$ elem. parami różnych

$$P[\{X_1, .., X_k\} = A] = \frac{1}{\binom{n}{k}}$$

$A \subseteq \{1..n\}$

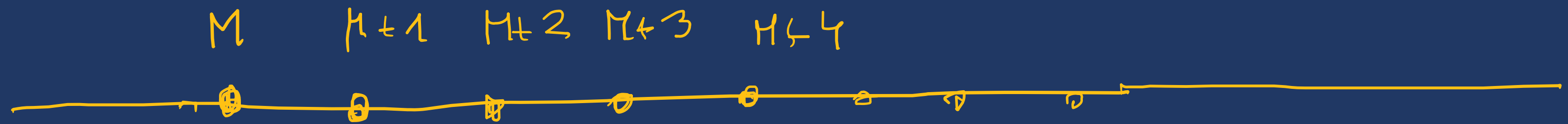
- przeczytaj pierwsze k elem; wstaw do $[X_1 .. X_k]$
- dla $n > k$.

```
if (random(0,1) < k/n) {
    i = random({1,..,k});
    X_i = (n, x)
}
```



POPRAWNOŚĆ!!!

ZADANIE



zmiana

L : taka liczba, że w momencie M+L reszta zmieniona

$$L \in \{1, 2, 3, \dots\}$$

$$P[L \geq 1] = 1$$

$$P[L \geq 2] = 1 - \frac{1}{M+1} = \frac{M}{M+1}$$

$$P[L > 2] = P[L \geq 3] = \left(1 - \frac{1}{M+1}\right)\left(1 - \frac{1}{M+2}\right) = \frac{M}{M+1} \cdot \frac{M+1}{M+2}$$

$$P[L > k] = \frac{M}{\cancel{M+1}} \cdot \frac{\cancel{M+1}}{\cancel{M+2}} \cdot \ldots \cdot \frac{\cancel{M+(k-1)}}{M+k} = \frac{M}{M+k}$$

$$P[L \leq k] = 1 - \frac{M}{M+k} = \frac{k}{M+k}$$

(*) $P[L \le k] = \dfrac{k}{M+k}$!!!

$= \dfrac{i}{k+i} - \dfrac{(i-1)}{M+(i-1)}$

$P_i = P[L=i]$

$P[L \le k] = P_1 + .. + P_k$ — u

$E[L] = \sum_{i=1} i\, p_i = \sum_{k} \frac{k}{M+k} = \infty$

zmienną o rozkł. (*)

1) $u \in$ uniform $(0,1)$ [u = random $(0,1)$]

2) znajdź k t. że

$p_1 + .. + p_{k-1} < u \le p_1 + .. + p_k$

zwróć k

nierówność Markowa

$\sum a_n \leftarrow$ zbieżny

$\rightarrow \lim a_n = 0$

szukamy <u>min</u> k t. że

$$P[L \leq k] \geqslant u$$

$$\frac{K}{M+K} \geqslant u$$

$$K \geqslant Mu + K \cdot u$$

$$K(1-u) \geqslant Mu$$

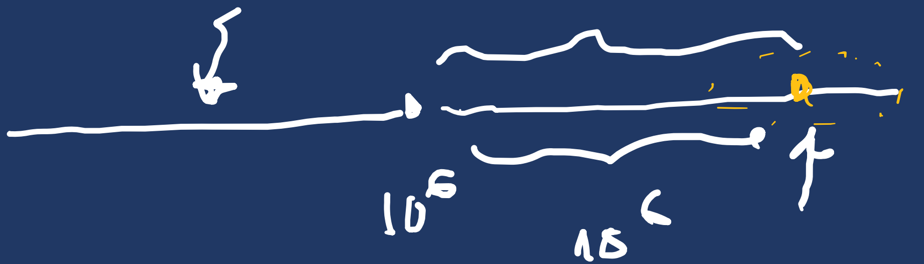
$$K \geqslant \frac{M \cdot u}{1-u}$$

$$K = \left\lceil \frac{M \cdot u}{1-u} \right\rceil$$

Modyfikacja alg. R:

```
n = 0;  c = 0;  data = nil;  p = 0;  L = 1;
on Read(x) {
    n++;  c++;
    if (c == L) {
        data = x;
        p    = n;
        u    = random(0,1)
        L    = ceiling((n*u)/(1-u));
        c    = 0;
    }
}
```

p $pr[\text{zmiany}] \approx \frac{1}{10^6}$



"catch" eliminacji losowości"

ZAŁ: znamy ogr. na całk. dług. strumienia

N (np. $N = 10^{16}$)

Przed uruchomieniem "obserwatora":

generujemy momenty zmian

$L_1, L_2, L_3, L_4, \dots, L_k, \dots$

$L_1 = 1$

dziękuję dr M. Gębali za korektę

$$L_{k+1} = L_k + \left\lceil \frac{L_k \cdot u}{1-u} \right\rceil$$

t. że $L_k \geq N$.