

MEDIAN TRICK

APPLICHIAMO LA MEDIAN TRICK AL PROBABILISTIC COUNTING PER AVERE DUE PARAMETRI INDIPENDENTI PER L'ENORE DI CAMPIONAMENTO E PER LA PROB. DI FALLIMENTO.

PER PROBABILISTIC COUNTING SAPPIAMO CHE

$$\hat{F} \in \left[\frac{F_0}{c}, cF_0 \right] \text{ CON PROB. } 1 - \frac{\delta}{c} \quad \forall c \geq 2$$

VUOLIAMO MODIFICARE L'ALGORITMO IN MODO CHE

$$\hat{F} \in \left[\frac{F_0}{c}, cF_0 \right] \text{ CON PROB. } 1 - \delta \quad \boxed{\begin{array}{l} \forall c \geq 2 \\ \delta \in (0,1) \end{array}}$$

MEDIAN TRICK

→ IL VALORE DI n VERRÀ DEFINITO DOPO

- ESEGUIAMO K ISTANZE DI PROB. COUNTING
 $\hookrightarrow$ OGNUNA CON DIVERSE FUNZIONI DI HASH.
- SIA 2^{R_i} LA STIMA DELLA i -ESIMA RIPETIZIONE
- LA STIMA FINALE È DATA DA

$$\hat{F} = \text{MEDIANA} \{ 2^{R_0}, 2^{R_1}, \dots, 2^{R_{K-1}} \}$$

(5)

TEOREMA: Sia $0 < \delta \in (0, 1)$. Se $K = \frac{6C}{C-2} \ln(1/2\delta)$ ALLORA ABBIAMO

$$\Pr \left[F \in \left[\frac{F_0}{C}, CF_0 \right] \right] \geq 1 - \delta$$

PROVA: Se $F > CF_0$, VUOL DIRE CHE ALMENO $\frac{K}{2}$

ISTANZE HANNO SOVRASTIMATO F .

Sia $X_i = \begin{cases} 1 & \text{se istanza } i \text{ sovrastima } (F_i > CF_0) \\ 0 & \text{altrimenti,} \end{cases}$

$$\Pr[X_i] = \frac{1}{C}, \quad E[X_i] = 1/C$$

$$E[X] = \frac{K}{C}$$

Sia $X \equiv$ NUMERO DI ISTANZE CHE SOVRASTIMANO. $X = \sum_{i=1}^{n-1} X_i$

$$\Pr \left[F > CF_0 \right] = \Pr \left[X \geq \frac{K}{2} \right] = \left(\text{poiché } E[X] = \frac{K}{C} \right) =$$

$$= \Pr \left[X \geq (1 + \varepsilon) E[X] \right] \quad \text{con } \varepsilon = \frac{C}{2} - 1 \geq 1$$

$$\leq \text{CHERNOFF BOUND} \leq e^{-\frac{\varepsilon^2 E[X]}{3}} = e^{-\frac{C-2}{6C} K} \leq \delta/2$$

con $\varepsilon \geq 1$

$$\text{PERCHÉ } K = \frac{6C}{C-2} \ln(1/2\delta)$$

SIMILMENTE SI PUÒ DIMOSTRARE CHE

$$\Pr \left[F < \frac{F_0}{C} \right] \leq \delta/2.$$

$$\text{QUINDI } \Pr \left[F \in \left[\frac{F_0}{C}, CF_0 \right] \right] \geq 1 - \delta.$$



6

IL TEOREMA SI PUÒ DIMOSTRARE ANCHE QUANDO

$C \in [2, 4]$. IN QUESTO CASO SI USA IL

CHEBYSHEV BOUND:

$$P_2 [x < (1+\epsilon) E[x]] \leq e^{-\frac{\epsilon^2 n}{2}} \text{ con } \epsilon \in [0, 1]$$

$$\text{PER } C \in [2, 4], \quad K = \frac{8}{(C-2)^2} \ln(1/25).$$