Randomized algo's with prob. of error

Las Vegas algo: Always correct, probably fast

Monte Carlo algo: Always fast, probably correct

\[ \text{Area} = 1 \]

\[ \text{Pr}(p \in B) = \text{Area}(B) \]

Est. area = Fraction of samples in B

Monte Carlo: Error rate is confidence 1 - 8

\[ \text{Pr}[(\text{wrong})/\text{bad}] < 8 \]

Set membership: Insert Query with error 8

One-sided error

\[ x \in S \Rightarrow \text{YES} \]

\[ x \notin S \Rightarrow \text{NO with prob} 1 - 8 \]

YES w/ prob < 8
ZATOCODING APPLIED TO MECHANICAL ORGANIZATION OF KNOWLEDGE

Simultaneously Selective Patterns

flash 17 23 34 55

camera 1 8 29 54

selective device 5 11 15 59

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Zatocodes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>selective device</td>
<td>5 11 15 59</td>
<td>U. S. Patent No. 2,095,000</td>
</tr>
<tr>
<td>film tally</td>
<td>14 17 22 50</td>
<td>Rapid Selector-Calculator</td>
</tr>
<tr>
<td>photo-electric sensing</td>
<td>1 11 54 40</td>
<td>Richard S. Morse, Rochester, N. Y.</td>
</tr>
<tr>
<td>audio frequency code</td>
<td>9 16 29 51</td>
<td></td>
</tr>
<tr>
<td>camera</td>
<td>1 8 29 54</td>
<td>one claim</td>
</tr>
<tr>
<td>flash</td>
<td>17 23 54 59</td>
<td></td>
</tr>
<tr>
<td>counting</td>
<td>8 26 55 57</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. ZATOCODING, illustrated with a 5 by 8 inch edge notched Zatocard for the tally. Note that the random Zatocode patterns in the field overlap and intermingle. Selection on the combination of three descriptors, "flash," "camera," and "selective device," is according to the inclusion of the pattern of arrows into the pattern of notches in the coding field. Zatocards are sorted by the selector shown in Figure 2.
Bloom Filter: bits $B[0..m-1]$

+ $k$ hash functions $h_i: U \rightarrow [m]$

**ideal random**

$\text{Insert}(x) :$

\begin{align*}
\text{for } i = 1 \text{ to } k \\
B[h_i(x)] &= -1
\end{align*}$

$\text{Member?}(x) :$

\begin{align*}
\text{for } i = 1 \text{ to } k \\
\text{if } B[h_i(x)] = 0 \\
\text{false} \\
\text{true}
\end{align*}$

$\Pr[h_i(x) = j] = \frac{1}{m}$

$\Pr(\text{Insert(x) does not set } B[j] = 1) = \left(1 - \frac{1}{m}\right)^k$

After $n$ Inserts:

$\Pr[\text{not set } B[j] = 1] = \left(1 - \frac{1}{m}\right)^kn < e^{-kn/m}$

$\text{WMU} \neq 0$

$e^x \geq 1 + x$

$\Pr[\text{false positive}] = (1-p)^k$

$\delta = (1-e^{-kn/m})^k$ (sort of)
\[
\ln S = \ln (1-p)^k = k \ln (1-p) = \frac{m}{n} \ln p \ln (1-p)
\]

\[\max \text{ at } p = \frac{1}{2} \Rightarrow k = \frac{m}{n} \cdot \ln 2\]

\[S = \left(\frac{1}{2}\right)^{m/n} \approx 0.6185^{m/n}\]

\[m = \frac{\log (1/S)}{\ln 2} \cdot n \Rightarrow \text{error prob } \delta\]

\[= O(n \cdot \log (1/S))\]

\[S = 1\% \Rightarrow 10n \text{ bits } \quad k = 7\]

\[m = 32n \Rightarrow k = 22 \quad \delta \approx 2 \cdot 10^{-7}\]

Mittenmacher
Broder
```
"How many times have I seen item x?"

\textit{Count Min Sketch} \hspace*{1cm} \text{wxd array of integers}
\hspace*{1cm} \text{with depth}\hspace*{1cm} \text{universal}
\hspace*{1cm} \text{d hash functions}
\hspace*{1cm} \text{h}_i : \mathbb{N} \rightarrow [w]

\textbf{Process}(x):
\begin{align*}
\text{For } i \leq 1 \text{ to d} \\
\text{Count}[i, h_i(x)]++
\end{align*}

\textbf{Estimate}(x):
\[ \min_i \text{ Count}[i, h_i(x)] \]

If we choose w and d correctly
\[ \Pr[\text{Estimate}(x) > \text{freq}(x) + \beta \cdot N] \leq \delta \]
```
\[ w = \left\lceil \frac{e}{\varepsilon} \right\rceil \quad d = \left\lceil \ln \left( \frac{1}{\varepsilon} \right) \right\rceil \]

\[ E[X_{i,x}] = \# \text{collisions with } x \text{ in row } i \]
\[ = \sum_{y \neq x} p[h_i(x) = h_i(y)] \cdot \text{Freq}(y) \]
\[ \leq \frac{N}{w} \]

Markov's inequality.

\[ \Pr \left[ X_{i,x} > \varepsilon N \right] \leq \frac{1}{we} \]

\[ \Pr \left[ \text{Est} > \text{Freq}(x) + \varepsilon N \right] \leq \left( \frac{1}{we} \right)^d < \delta \]