Big Data Management and Analytics
Assignment 7

Assignment 7-1

Exponential Histograms

Purpose: solve the problem of counting number of x within a sliding window of size N

Given:

Sequence S = (x, x, o, x, o, o, x, x, x, x, o, x, x, o, x, x)

Window size $N = 8$

Error parameter $\epsilon = \frac{1}{2}$

Assignment 7-1

Sequence S = (x, x, o, x, o, o, x, x, x, x, o, x, x, o, x, x)

Window size $N = 8$

Error parameter $\epsilon = \frac{1}{2}$

Max. # of buckets of same size $\tau = \frac{\left\lceil \frac{1}{\epsilon} \right\rceil}{2} + 2 = 3$

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	*# buckets of same size* $= \tau$?
1	1_1	x	1	0	no
2	$1_1, 1_2$	x	2	0	no
3	$1_1, 1_2$	o	2	0	no
4	$1_1, 1_2, 1_4$ $\rightarrow 2_2, 1_4$	x	3	2	yes

Merge two oldest buckets of same size with the largest timestamp of both buckets!

Assignment 7-1

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	$\# \textit{ buckets of same size} = \tau$?
1	1_1	x	1	0	no
2	$1_1,\ 1_2$	x	2	0	no
3	$1_1,\ 1_2$	o	2	0	no
4	$2_2, 1_4$	x	3	2	yes
5	$2_2, 1_4$	o	3	2	no
6	$2_2, 1_4$	o	3	2	no
7	$2_2, 1_4,\ 1_7$	x	4	2	no
8	$2_2, 1_4,\ 1_7,\ 1_8$ → $2_2, 2_7,\ 1_8$	x	5	2	Yes
9	$2_2, 2_7,\ 1_8,\ 1_9$	x	6	2	no

Merge two oldest buckets of same size with the largest timestamp of both buckets!

Assignment 7-1

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	# *buckets of same size* $= \tau$?
10	$2_2, 2_7, 1_8, 1_9, 1_{10}$ → $2_7, 1_8, 1_9, 1_{10}$ → $2_7, 2_9, 1_{10}$	x	7 7-2=5	2	yes

Merge two oldest buckets of same size with the largest timestamp of both buckets!

3. $b_l := b_{l-1} \rightarrow 2_7$
$\text{LAST} = b_l.size \rightarrow 2$

1. $TOTAL = TOTAL - b_l.size \rightarrow 7 - 2 = 5$

2. Oldest timestamp $t_l \leq t_i - N \rightarrow 2 \leq 10 - 8$
drop the oldest bucket 2_2

Assignment 7-1

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	$\# \, buckets \, of \, same \, size = \tau$?
10	$2_7, 2_9, 1_{10}$	x	5	2	yes
11	$2_7, 2_9, 1_{10}$	o	5	2	no
12	$2_7, 2_9, 1_{10}, 1_{12}$	x	6	2	no
13	$2_7, 2_9, 1_{10}, 1_{12}, 1_{13}$ → $2_7, 2_9, 2_{12}, 1_{13}$ → $4_9, 2_{12}, 1_{13}$	x	7	4	yes

Merge two oldest buckets of same size with the largest timestamp of both buckets!

Last bucket was merged!
$LAST := size \; of \; the \; new \; created \; last \; bucket = 4$

Merge two oldest buckets of same size with the largest timestamp of both buckets!

Assignment 7-1

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	$\# \, buckets$ $of \, same \, size = \tau$?
13	$4_9, 2_{12}, 1_{13}$	x	7	4	yes
14	$4_9, 2_{12}, 1_{13}$	o	7	4	no
15	$4_9, 2_{12}, 1_{13}, 1_{15}$	x	8	4	no
16	$4_9, 2_{12}, 1_{13}, 1_{15}, 1_{16}$ → $4_9, 2_{12}, 2_{15}, 1_{16}$	x	9	4	yes

Assignment 7-1

Timest. t_i	Buckets b_i	Element x_i	TOTAL	LAST	*# buckets of same size* $= \tau$?
13	$4_9, 2_{12}, 1_{13}$	x	7	4	yes

Estimating total number of x's within the sliding window of size 8 in the exponential histogram:

x's = EH.TOTAL – EH.LAST/2 = 7 - 4/2 = 5

Sequence S = (x, x, o, x, o, o, x, x, x, x, o, x, x, o, x, x)

Exact number of x's in sliding window [6:13] : 6

Assignment 7-2

Hoeffding Trees

Core idea: For choosing the best split attribute for a node, a small subset of the training examples may suffice

Question: How many instances are required?

Solution: Utilize the Hoeffding bound

Assignment 7-2

Given:

8 examples of drivers with the attributes:

- Time since getting the driving license (1-2 years, 2-7 years, > 7 years)
- Gender (female, male)
- Residential area (urban, rural)

Further: $\delta = 0.2, \quad N_{min} = 2$

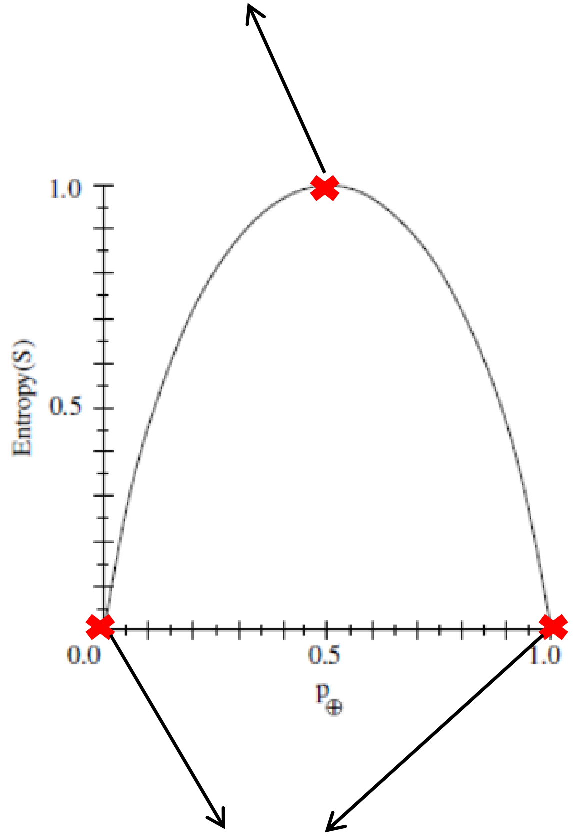
- Use information gain
- Output is nominal risk class → two attributes, R=1

Assignment 7-2: RECAP

$$entropy(D_i) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

$$IG(D_i, A) = entropy(D_i) - \sum_{v \in Values(A)} \frac{|D_{i_v}|}{|D_i|} entropy(D_{i_v})$$

If there is an equal number of positive and negative examples



If all members belong to the same class

Assignment 7-2

Person	Time since license	Gender	Area	Risk class
1	1-2	m	urban	low
2	2-7	m	rural	high
3	>7	f	rural	low
4	1-2	f	rural	high
5	>7	m	rural	high
6	1-2	m	rural	high
7	2-7	f	urban	low
8	2-7	m	urban	low

Assignment 7-2

$$\varepsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n}} = \sqrt{\frac{1 \ln(1/\delta)}{2n}} = \sqrt{\ln(1/\delta)} \cdot \sqrt{1/(2n)}$$

Where

- Confidence $\boldsymbol{\delta}$: what probability do we allow of 'failure'? (How much do we accept a deviation $> \varepsilon$)
- Range **R**: e.g. a probability range from 0 to 1
- # of training examples **n**
- Accuracy $\boldsymbol{\varepsilon}$: How much do we want to allow the empirical mean to differ from the true mean

For $n = 2,4,6,8$ this yields:

$$\varepsilon_2 \approx 0.634$$
$$\varepsilon_4 \approx 0.448$$
$$\varepsilon_6 \approx 0.366$$
$$\varepsilon_8 \approx 0.317$$

Assignment 7-2

We initialize an empty tree. Now insert the first two records:

Person	Time since license	Gender	Area	Risk class
1	1-2	m	urban	low
2	2-7	m	rural	high

1. Compute the entropy for the first two records:
   $entropy(D_2) = -\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right) =$ 1, due to the fact that both classes have a probability of 50%

2. Compute the information gain (IG) for all three attributes (time, gender, area):
   $IG(time, D_2) = entropy(D_2) - (0.5 \text{ entropy}(D_2|t = 1-2) + 0.5 \text{ entropy } (D_2|t = 2-7) + 0\ entropy(D_2|t > 7)) = 1 - \big(0.5 * \left(-0 * log_2(0) - 1 * log_2(1)\right) + 0.5 * log_2(1) - 0 * log_2(0)\big) + 0.5\ * \left(-0 * log_2(0) - 0 * log_2(0)\right)\big) = 1 - (0 + 0 + 0) = 1$

Assignment 7-2

2. $IG(gender, D_2) = entropy(D_2) - (1 \text{ entropy}(D_2|g = m) + 0 \text{ entropy } (D_2|g = f)) = 1 - (1 * (-\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right)) + 0 * \left(-0 * log_2(0) - 0 * log_2(0)\right)) = 1 - (1 + 0) = 0$

   $IG(area, D_2) = entropy(D_2) - (0.5 \text{ entropy}(D_2|a = u) + 0.5 \text{ entropy } (D_2|a = r)) = 1 - (0.5 * \left(-0 * log_2(0) - 1 * log_2(1)\right) + 0.5 * \left(-1 * log_2(1) - 0 * log_2(0)\right)) = 1 - (0 + 0) = 1$

Assignment 7-2

3. Compare the best with the second best result:
   $IG(time, D_2) - IG(area, D_2) = 1 - 1 = 0 < \varepsilon_2 \approx 0.634$
   → continue with more samples!

Assignment 7-2

Now take two more records (the first four records)

Person	Time since license	Gender	Area	Risk class
1	1-2	m	urban	low
2	2-7	m	rural	high
3	>7	f	rural	low
4	1-2	f	rural	high

Assignment 7-2

Again, proceed as follows:

1. Compute the entropy for the first two records:
$entropy(D_4) = -\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right) =$ 1, due to the fact that both classes have a probability of 50%

2. Compute the information gain (IG) for all three attributes (time, gender, area):
$IG(time, D_4) = entropy(D_4) - (\frac{2}{4}\text{ entropy}(D_4|t = 1 - 2) + \frac{1}{4}\text{ entropy }(D_4|t = 2 - 7) + \frac{1}{4}\ entropy(D_4|t > 7)) = 1 - \left(\frac{2}{4} * \left(-\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right)\right) + \frac{1}{4} * \left(-1 * log_2(1) - 0 * log_2(0)\right) + \frac{1}{4} * \left(-0 * log_2(0) - 1 * log_2(1)\right)\right) = 1 - \left(\frac{2}{4} * 1 + 0 + 0\right)$
$= 0.5$

Assignment 7-2

Again, proceed as follows:

2. $IG(gender, D_4) = entropy(D_4) - (\frac{2}{4}\text{ entropy}(D_4|g = m) + \frac{2}{4}\text{ entropy }(D_4|g = f)) = 1 - (\frac{2}{4} * (-\frac{1}{2} * log_2(\frac{1}{2}) - \frac{1}{2} * log_2(\frac{1}{2})) + \frac{2}{4} * (-\frac{1}{2} * log_2(\frac{1}{2}) - \frac{1}{2} * log_2(\frac{1}{2}))) = 1 - \left(\frac{1}{2} * 1 + \frac{1}{2} * 1\right) = 0$

   $IG(area, D_4) = entropy(D_4) - (\frac{1}{4}\text{entropy}(D_4|a = u) + \frac{3}{4}\text{ entropy }(D_4|a = r)) \approx 1 - (\frac{1}{4} * (-0 * log_2(0) - 1 * log_2(1)) + \frac{3}{4} * (-\frac{2}{3} * log_2(\frac{2}{3}) - \frac{1}{3} * log_2(\frac{1}{3}))) \approx 1 - \left(0 + \frac{3}{4} * 0.918\right) \approx 0.311$

Assignment 7-2

3. Compare the best with the second best result:
   $IG(time, D_4) - IG(area, D_4) = 0.5 - 0.311 \approx 0.189 < \varepsilon_4 \approx 0.448$
   → continue with more samples!

Assignment 7-2

Now take two more records (the first six records)

Person	Time since license	Gender	Area	Risk class
1	1-2	m	urban	low
2	2-7	m	rural	high
3	>7	f	rural	low
4	1-2	f	rural	high
5	>7	m	rural	high
6	1-2	m	rural	high

Assignment 7-2

Again, proceed as follows:

1. Compute the entropy for the first six records:

$$entropy(D_6) = -\frac{2}{3} log_2\left(\frac{2}{3}\right) - \frac{1}{3} log_2\left(\frac{1}{3}\right) \approx 0.918$$

2. Compute the information gain (IG) for all three attributes (time, gender, area):

$IG(time, D_6) = entropy(D_6) - (\frac{3}{6}\text{entropy}(D_6|t = 1 - 2) + \frac{1}{6}\text{ entropy }(D_6|t = 2 - 7) + \frac{2}{6}\ entropy(D_6|t > 7)) \approx 0.918\ - (\frac{3}{6} * (-\frac{2}{3} * log_2(\frac{2}{3}) - \frac{1}{3} * log_2(\frac{1}{3})) + \frac{1}{6} * (-1 * log_2(1) - 0 * log_2(0))\ + \frac{2}{6} * (-\frac{1}{2} * log_2(\frac{1}{2}) - \frac{1}{2} * log_2(\frac{1}{2})))\ \approx 0.918 - \left(\frac{3}{6} * 0.918 + 0 + \frac{2}{6} * 1\right) \approx 0.126$

Assignment 7-2

Again, proceed as follows:

2. $IG(gender, D_6) = entropy(D_6) - (\frac{4}{6}\text{entropy}(D_6|g = m) + \frac{2}{6}\text{ entropy }(D_6|g = f)) \approx$
$0.918 - (\frac{4}{6} * (-\frac{3}{4} * log_2(\frac{3}{4}) - \frac{1}{4} * log_2(\frac{1}{4})) + \frac{2}{6} * (-\frac{1}{2} * log_2(\frac{1}{2}) - \frac{1}{2} * log_2(\frac{1}{2}))) \approx$
$0.918 - \left(\frac{4}{6} * 0.811 + \frac{2}{6} * 1\right) \approx 0.044$

$IG(area, D_6) = entropy(D_6) - (\frac{1}{6}\text{entropy}(D_6|a = u) + \frac{5}{6}\text{entropy }(D_6|a = r)) \approx$
$0.918 - (\frac{1}{6} * (-0 * log_2(0) - 1 * log_2(1)) + \frac{5}{6} * (-\frac{4}{5} * log_2(\frac{4}{5}) - \frac{1}{5} * log_2(\frac{1}{5}))) \approx$
$0.918 - \left(0 + \frac{5}{6} * 0.722\right) \approx 0.316$

Assignment 7-2

3. Compare the best with the second best result:
   $IG(area, D_6) - IG(time, D_6) = 0.316 - 0.126 \approx 0.19 < \varepsilon_6 \approx 0.366$
   → continue with more samples!

Assignment 7-2

Now take two more records (all eight records)

Person	Time since license	Gender	Area	Risk class
1	1-2	m	urban	low
2	2-7	m	rural	high
3	>7	f	rural	low
4	1-2	f	rural	high
5	>7	m	rural	high
6	1-2	m	rural	high
7	2-7	f	urban	low
8	2-7	m	urban	low

Assignment 7-2

Again, proceed as follows:

1. Compute the entropy for all eight records:

$$entropy(D_8) = -\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right) = 1$$

2. Compute the information gain (IG) for all three attributes (time, gender, area):

$IG(time, D_8) = entropy(D_8) - (\frac{3}{8}\text{entropy}(D_8|t = 1 - 2) + \frac{3}{8}\text{ entropy }(D_8|t = 2 - 7) + \frac{2}{8} entropy(D_8|t > 7)) \approx 1 - \left(\frac{3}{8} * (-\frac{2}{3} * log_2\left(\frac{2}{3}\right) - \frac{1}{3} * log_2\left(\frac{1}{3}\right)) + \frac{3}{8} * (-\frac{1}{3} * log_2\left(\frac{1}{3}\right) - \frac{2}{3} * log_2\left(\frac{2}{3}\right)) + \frac{2}{8} * (-\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right))\right) \approx 1 - \left(\frac{3}{8} * 0.918 + \frac{3}{8} * 0.918 + \frac{2}{8} * 1\right) \approx 0.065$

Assignment 7-2

Again, proceed as follows:

2. $IG(gender, D_8) = entropy(D_8) - (\frac{5}{8}\text{entropy}(D_8|g = m) + \frac{3}{8}\text{ entropy }(D_8|g = f)) \approx 1 - \left(\frac{5}{8} * \left(-\frac{3}{5} * log_2\left(\frac{3}{5}\right) - \frac{2}{5} * log_2\left(\frac{2}{5}\right)\right) + \frac{3}{8} * \left(-\frac{1}{3} * log_2\left(\frac{1}{3}\right) - \frac{2}{3} * log_2\left(\frac{2}{3}\right)\right)\right) \approx 1 - \left(\frac{5}{8} * 0.971 + \frac{3}{8} * 0.918\right) \approx 0.049$

   $IG(area, D_8) = entropy(D_8) - (\frac{3}{8}\text{entropy}(D_8|a = u) + \frac{5}{8}\text{entropy }(D_8|a = r)) \approx 1 - \left(\frac{3}{8} * (-0 * log_2(0) - 1 * log_2(1)) + \frac{5}{8} * (-\frac{4}{5} * log_2\left(\frac{4}{5}\right) - \frac{1}{5} * log_2\left(\frac{1}{5}\right))\right) \approx 1 - \left(0 + \frac{5}{8} * 0.722\right) \approx 0.549$

Assignment 7-2

3. Compare the best with the second best result:
   $IG(area, D_8) - IG(time, D_8) = 0.549 - 0.065 \approx 0.484 > \varepsilon_8 \approx 0.317$
   → split at 'area' attribute!

   → New leafs are empty and have no 'area' attribute.
   → Further splits are not required until new data arrives.

Assignment 7-2

Computing the value of δ at which the tree would still consist only of the leaf:

The minimal ε for which a further split would be required: 0.484

$$\varepsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n}} \Rightarrow 2n\varepsilon^2 = \ln(1/\delta) \Rightarrow \delta = \frac{1}{\exp(2n\varepsilon^2)} \approx \frac{1}{\exp(16 * 0.484^2)} \approx 0.0235$$

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 0$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 0$$

elements of D
(e, f, Δ)

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    e_i ≔ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 1$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 0$$

elements of D
(e, f, Δ)

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    e_i := next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \; mod \; \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \lceil \frac{1}{\epsilon} \rceil = 4$$

$$N = 1$$

$$b_{curr} = \lceil \frac{N}{\omega} \rceil = 1$$

elements of D
(e, f, Δ)

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \lceil \frac{1}{\epsilon} \rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \lceil \frac{N}{\omega} \rceil$ ←
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 1$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$

$A \notin D$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$ ←
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 1$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 2$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 2$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    e_i := next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 2$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1{,}0)$
$(B, 1{,}0)$

$B \notin D$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$ ←
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 2$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1,0)$
$(B, 1,0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ⟵
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 3$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1,0)$
$(B, 1,0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 3$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1,0)$
$(B, 1,0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 3$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$
$(B, 1, 0)$
$(C, 1, 0)$

$C \notin D$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$ ←
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \le b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 3$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1{,}0)$
$(B, 1{,}0)$
$(C, 1{,}0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    e_i ≔ next object from S ←
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1,0)$
$(B, 1,0)$
$(C, 1,0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
$D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
while S **do**
$e_i :=$ next object from S
$N \mathrel{+}= 1$
$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
if $e_i \in D$ **then**
increment e_i's frequency by 1
else
$D.add((e_i, 1, b_{curr} - 1))$
whenever $N \equiv 0 \bmod \omega$ **do**
foreach entry (e, f, Δ) in D **do**
if $f + \Delta \leq b_{curr}$ **then**
delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1{,}0)$
$(B, 1{,}0)$
$(C, 1{,}0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1{,}0)$
$(B, 1{,}0)$
$(C, 2{,}0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(A, 1, 0)$
$(B, 1, 0)$
$(C, 2, 0)$

$$1 + 0 = 1 \ \leq 1$$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i \coloneqq$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(B, 1, 0)$
$(C, 2, 0)$

$1 + 0 = 1 \leq 1$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(C, 2, 0)$

$2 + 0 = 2 > 1$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 4$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(C, 2, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
 $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
 while S **do**
  $e_i :=$ next object from S
  $N \mathrel{+}= 1$
  $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
  if $e_i \in D$ **then**
   increment e_i's frequency by 1
  else
   $D.add((e_i, 1, b_{curr} - 1))$
  whenever $N \equiv 0 \bmod \omega$ **do**
   foreach entry (e, f, Δ) in D **do**
    if $f + \Delta \leq b_{curr}$ **then**
     delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 5$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 1$$

elements of D (e, f, Δ)
$(C, 2, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$ ⟵
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 5$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 2, 0)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$ ←
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 5$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 2, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$ ⬅
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 5$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 2, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 6$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 2, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 6$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 2, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$ ⬅
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \le b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 6$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3,0)$
$(A, 1,1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 6$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3,0)$
$(A, 1,1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 7$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 7$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$ ⬅
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 7$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3,0)$
$(A, 1,1)$
$(B, 1,1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$ ⟵
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 7$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3{,}0)$
$(A, 1{,}1)$
$(B, 1{,}1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ⬅
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3,0)$
$(A, 1,1)$
$(B, 1,1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 1, 1)$
$(B, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 2, 1)$
$(B, 1, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1 ←
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3{,}0)$
$(A, 2{,}1)$
$(B, 1{,}1)$

$3 + 0 = 3 > 2$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 2, 1)$
$(B, 1, 1)$

$$2 + 1 = 3 > 2$$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 2, 1)$
$(B, 1, 1)$

$$1 + 1 = 2 \leq 2$$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 8$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ←
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 9$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 2$$

elements of D (e, f, Δ)
$(C, 3, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \le b_{curr}$ **then**
          delete (e, f, Δ)
end

DATABASE SYSTEMS GROUP

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 9$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 3{,}0)$
$(A, 2{,}1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 9$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 4, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1 ←
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 9$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 4, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ←
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

DATABASE SYSTEMS GROUP

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 10$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 4, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$ ⬅
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 10$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 4, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$ ⬅
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 10$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 10$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ←
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 11$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \; mod \; \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 11$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 2, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset, N = 0, \omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    e_i := next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$ ←
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

DATABASE SYSTEMS GROUP

Assignment 7-3

LMU

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 11$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 3, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 11$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 3, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S ←
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 12$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5,0)$
$(A, 3,1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 12$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 5, 0)$
$(A, 3, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i \coloneqq$ next object from S
    $N \mathrel{+}= 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 12$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 6, 0)$
$(A, 3, 1)$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1 ←
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \bmod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \left\lceil \frac{1}{\epsilon} \right\rceil = 4$$

$$N = 12$$

$$b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil = 3$$

elements of D (e, f, Δ)
$(C, 6, 0)$
$(A, 3, 1)$

$$6 + 0 = 6 > 3$$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \left\lceil \frac{1}{\epsilon} \right\rceil$
  while S **do**
    $e_i :=$ next object from S
    $N += 1$
    $b_{curr} = \left\lceil \frac{N}{\omega} \right\rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end

Assignment 7-3

Data stream S

Time t	1	2	3	4	5	6	7	8	9	10	11	12
Item e	A	B	C	C	A	C	B	A	C	C	A	C

Error threshold $\epsilon = 0.25$

$$\omega = \lceil \frac{1}{\epsilon} \rceil = 4$$

$$N = 12$$

$$b_{curr} = \lceil \frac{N}{\omega} \rceil = 3$$

elements of D (e, f, Δ)
$(C, 6,0)$
$(A, 3,1)$

$3 + 1 = 4 > 3$

Algorithm LossyCounting
Input: data stream S, error threshold ϵ
begin
  $D = \emptyset$, $N = 0$, $\omega = \lceil \frac{1}{\epsilon} \rceil$
  while S **do**
    $e_i :=$ next object from S
    N += 1
    $b_{curr} = \lceil \frac{N}{\omega} \rceil$
    if $e_i \in D$ **then**
      increment e_i's frequency by 1
    else
      $D.add((e_i, 1, b_{curr} - 1))$
    whenever $N \equiv 0 \mod \omega$ **do**
      foreach entry (e, f, Δ) in D **do**
        if $f + \Delta \leq b_{curr}$ **then**
          delete (e, f, Δ)
end