Random sampling in geometry

- We have n objects in the set N, and subsets of N can be the *defining* elements of *configurations* comprising the set $\Pi = \Pi(N)$ of configurations.

Random sampling in geometry

- We have n objects in the set N, and subsets of N can be the *defining* elements of *configurations* comprising the set $\Pi = \Pi(N)$ of configurations.
- Let $\sigma \in \Pi$ be one such configuration.

Random sampling in geometry

- We have n objects in the set N, and subsets of N can be the *defining* elements of *configurations* comprising the set $\Pi = \Pi(N)$ of configurations.
- Let $\sigma \in \Pi$ be one such configuration.
- Imagine a one-dimensional space (the real line) with n distinct points and the pairs of n points as configurations, which are actually linear intervals.

Random sampling in geometry

- We have n objects in the set N, and subsets of N can be the *defining* elements of *configurations* comprising the set $\Pi = \Pi(N)$ of configurations.
- Let $\sigma \in \Pi$ be one such configuration.
- Imagine a one-dimensional space (the real line) with n distinct points and the pairs of n points as configurations, which are actually linear intervals.
- If we fix a constant $r < n$, we may take a *random sample* R of r elements, selected out of the n elements in N.

Random sampling in geometry

- Sampling is done without repetitions; each time an element is selected independently and randomly.

Random sampling in geometry

- Sampling is done without repetitions; each time an element is selected independently and randomly.
- Here, the cardinality $d(\sigma)$ of the seti $D(\sigma)$ of *triggers* or *defining elements* of any of these configurations σ is exactly 2.

Random sampling in geometry

- Sampling is done without repetitions; each time an element is selected independently and randomly.
- Here, the cardinality $d(\sigma)$ of the seti $D(\sigma)$ of *triggers* or *defining elements* of any of these configurations σ is exactly 2.
- Each configuration σ may contain (intersect) a set $L(\sigma)$ of elements from N called *stoppers*. The number of stoppers is denoted by $l(\sigma)$ and is called the *conflict size*.

Random sampling in geometry

- Sampling is done without repetitions; each time an element is selected independently and randomly.
- Here, the cardinality $d(\sigma)$ of the seti $D(\sigma)$ of *triggers* or *defining elements* of any of these configurations σ is exactly 2.
- Each configuration σ may contain (intersect) a set $L(\sigma)$ of elements from N called *stoppers*. The number of stoppers is denoted by $l(\sigma)$ and is called the *conflict size*.
- We say that $\sigma \in \Pi$ is active over a subset $R \subseteq N$ if it occurs as an interval in $H(R)$, the partition formed on the line by R.

Random sampling in geometry

- Sampling is done without repetitions; each time an element is selected independently and randomly.
- Here, the cardinality $d(\sigma)$ of the seti $D(\sigma)$ of *triggers* or *defining elements* of any of these configurations σ is exactly 2.
- Each configuration σ may contain (intersect) a set $L(\sigma)$ of elements from N called *stoppers*. The number of stoppers is denoted by $l(\sigma)$ and is called the *conflict size*.
- We say that $\sigma \in \Pi$ is active over a subset $R \subseteq N$ if it occurs as an interval in $H(R)$, the partition formed on the line by R.
- This occurs if and only if R contains all the points defining σ but no point in conflict with σ.

Random sampling in geometry

- Such configurations are called *active configurations* of Π over the random sample R.

Random sampling in geometry

- Such configurations are called *active configurations* of Π over the random sample R.
- We show that every active configuration over the random sample R of cardinality r, would have conflict size $O(\frac{n}{r}\log r)$, with probability at least $\frac{1}{2}$.

Random sampling in geometry

- Such configurations are called *active configurations* of Π over the random sample R.
- We show that every active configuration over the random sample R of cardinality r, would have conflict size $O(\frac{n}{r}\log r)$, with probability at least $\frac{1}{2}$.
- Let $p(\sigma, r)$ denote the conditionall probability that R has no point in conflict with σ, given that R contains the points defining σ

$$p(\sigma, r) \leq (1 - \frac{l(\sigma)}{n})^{r-d(\sigma)} \qquad (.3)$$

Random sampling in geometry

- The intuitive justification is as follows. The interval being of conflict size $l(\sigma)$, the probability of choosing a conflicting point is at least $\frac{l(\sigma)}{n}$.

Random sampling in geometry

- The intuitive justification is as follows. The interval being of conflict size $l(\sigma)$, the probability of choosing a conflicting point is at least $\frac{l(\sigma)}{n}$.
- Since we select $r - d(\sigma)$ points without conflicts, the probability required is upper bounded as in Inequality .3.

Random sampling in geometry

- The intuitive justification is as follows. The interval being of conflict size $l(\sigma)$, the probability of choosing a conflicting point is at least $\frac{l(\sigma)}{n}$.
- Since we select $r - d(\sigma)$ points without conflicts, the probability required is upper bounded as in Inequality .3.
- However, since $1 - x \leq exp(-x)$ where $exp(x) = e^x$, we have

$$p(\sigma, r) \leq exp(-\frac{l(\sigma)}{n}(r - d(\sigma))) \qquad (.6)$$

Random sampling in geometry

- Since $d(\sigma) \leq 2$, putting $l(\sigma) \geq c(n \ln s)/(r-2)$ for some $c > 1$ and $s \geq r$, we get

$$p(\sigma, r) \leq exp(-c \ln s) = \frac{1}{s^c} \qquad (.7)$$

Random sampling in geometry

- Since $d(\sigma) \leq 2$, putting $l(\sigma) \geq c(n \ln s)/(r-2)$ for some $c > 1$ and $s \geq r$, we get

$$p(\sigma, r) \leq exp(-c \ln s) = \frac{1}{s^c} \qquad (.8)$$

- Now an active configuration σ due to the random sample R must be such that all its defining points must be in R.

Random sampling in geometry

- Since $d(\sigma) \leq 2$, putting $l(\sigma) \geq c(n \ln s)/(r-2)$ for some $c > 1$ and $s \geq r$, we get

$$p(\sigma, r) \leq exp(-c \ln s) = \frac{1}{s^c} \qquad (.9)$$

- Now an active configuration σ due to the random sample R must be such that all its defining points must be in R.
- In other words, $\sigma \in \Pi(R)$.

Random sampling in geometry

- Since $d(\sigma) \leq 2$, putting $l(\sigma) \geq c(n \ln s)/(r-2)$ for some $c > 1$ and $s \geq r$, we get

$$p(\sigma, r) \leq exp(-c \ln s) = \frac{1}{s^c} \qquad (.10)$$

- Now an active configuration σ due to the random sample R must be such that all its defining points must be in R.
- In other words, $\sigma \in \Pi(R)$.
- Let this probability be $q(\sigma, r)$.

Random sampling in geometry

- The probability that σ is an active configuration in the partition created by the random sample R is at most

$$p(\sigma, r)q(\sigma, r)$$

Random sampling in geometry

- The probability that σ is an active configuration in the partition created by the random sample R is at most

$$p(\sigma, r)q(\sigma, r)$$

- The probability that there is some active configuration created by the partition due to the random sample R with conflict size lower bounded by $\frac{cn \ln s}{r-2}$, is upper bounded by the sum of the probabilities over all such configurations

$$\sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} p(\sigma, r)q(\sigma, r)$$

Random sampling in geometry

- The probability that σ is an active configuration in the partition created by the random sample R is at most

$$p(\sigma, r)q(\sigma, r)$$

- The probability that there is some active configuration created by the partition due to the random sample R with conflict size lower bounded by $\frac{cn \ln s}{r-2}$, is upper bounded by the sum of the probabilities over all such configurations

$$\sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} p(\sigma, r)q(\sigma, r)$$

Random sampling in geometry

- $$\leq \sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} q(\sigma, r)/s^c \leq \frac{1}{s^c} \sum_{\sigma \in \Pi} q(\sigma, r)$$

Random sampling in geometry

- $$\leq \sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} q(\sigma, r)/s^c \leq \frac{1}{s^c} \sum_{\sigma \in \Pi} q(\sigma, r)$$

- Now the last summation in the above inequality is $E(\pi(R))$ and $\pi(R) = |\Pi(R)| = O(r^2)$. So, choosing $c > 2$ we can ensure that the probability of having a "long" active configuration in $\sigma \in \Pi(R)$ is less than $\frac{1}{2}$ for a random sample R.

The weak cutting lemma

- With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines.

The weak cutting lemma

- With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines.
- Any triangle that cuts k lines is divided into at most $2k^2$ parts.

The weak cutting lemma

- With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines.
- Any triangle that cuts k lines is divided into at most $2k^2$ parts.
- Suppose we have only t (arbitrary) triangles partitioning the whole plane containing the arrangement of n lines, and each such triangle is cut by at most k of the n given lines.

The weak cutting lemma

- With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines.
- Any triangle that cuts k lines is divided into at most $2k^2$ parts.
- Suppose we have only t (arbitrary) triangles partitioning the whole plane containing the arrangement of n lines, and each such triangle is cut by at most k of the n given lines.
- Since each of these at least $\frac{n^2}{2}$ cells has be be covered by only t triangles, each of which has at most $2k^2$ cells as stated above, we need to have $t \geq \frac{n^2}{4k^2} = \Omega(r^2)$ triangles, provided we fix $k \leq \frac{n}{r}$.

The weak cutting lemma

- With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines.
- Any triangle that cuts k lines is divided into at most $2k^2$ parts.
- Suppose we have only t (arbitrary) triangles partitioning the whole plane containing the arrangement of n lines, and each such triangle is cut by at most k of the n given lines.
- Since each of these at least $\frac{n^2}{2}$ cells has be be covered by only t triangles, each of which has at most $2k^2$ cells as stated above, we need to have $t \geq \frac{n^2}{4k^2} = \Omega(r^2)$ triangles, provided we fix $k \leq \frac{n}{r}$.

The weak cutting lemma: A $\frac{1}{r}$-cutting

- We show that a set of $O(r^2 \log^2 n)$ triangles can be used to ensure that less than $\frac{n}{r}$ lines of the arrangement of n lines cross each such triangle.

The weak cutting lemma: A $\frac{1}{r}$-cutting

- We show that a set of $O(r^2 \log^2 n)$ triangles can be used to ensure that less than $\frac{n}{r}$ lines of the arrangement of n lines cross each such triangle.
- Use a random sample $S \subset L$ of size $s = r \log n$ to create $O(s^2)$ regions as follows. If there are non-triangular regions, we triangulate them.

The weak cutting lemma: A $\frac{1}{r}$-cutting

- We show that a set of $O(r^2 \log^2 n)$ triangles can be used to ensure that less than $\frac{n}{r}$ lines of the arrangement of n lines cross each such triangle.
- Use a random sample $S \subset L$ of size $s = r \log n$ to create $O(s^2)$ regions as follows. If there are non-triangular regions, we triangulate them.
- A *bad* triangle T (defined by any three lines of the n lines in L) has strictly more than $k = \frac{n}{r}$ lines intersecting T.

The weak cutting lemma: A $\frac{1}{r}$-cutting

- We show that a set of $O(r^2 \log^2 n)$ triangles can be used to ensure that less than $\frac{n}{r}$ lines of the arrangement of n lines cross each such triangle.
- Use a random sample $S \subset L$ of size $s = r \log n$ to create $O(s^2)$ regions as follows. If there are non-triangular regions, we triangulate them.
- A *bad* triangle T (defined by any three lines of the n lines in L) has strictly more than $k = \frac{n}{r}$ lines intersecting T.
- Such a bad triangle is also called *interesting* if it appears in the triangulation of $S \subset L$, as one of the $O(s^2)$ triangles created by a random sample S of size s as mentioned above.

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.
- The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. Why?

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.
- The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. Why?
- So, the probability that some interesting triangle is bad is *strictly less* than unity, as we show now.

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.
- The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. Why?
- So, the probability that some interesting triangle is bad is *strictly less* than unity, as we show now.
- This probability is strictly less than $n^6(1 - \frac{k}{n})^s \leq n^6(1 - \frac{1}{r})^{6r \ln n} < n^6 e^{-6 \ln n} = n^6 n^{-6} = 1$

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.
- The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. Why?
- So, the probability that some interesting triangle is bad is *strictly less* than unity, as we show now.
- This probability is strictly less than $n^6(1 - \frac{k}{n})^s \leq n^6(1 - \frac{1}{r})^{6r \ln n} < n^6 e^{-6 \ln n} = n^6 n^{-6} = 1$
- Therefore, there exists a random sample S of size $s = 6r \log n$ such that the none of the $O(s^2)$ triangles induced by S meet more than $\frac{n}{r}$ lines of L.

The weak cutting lemma: A probabilistic argument

- There are at most n^6 interesting triangles since each triangle is defined by three of the $\binom{n}{2}$ points of intersections as vertices.
- The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. Why?
- So, the probability that some interesting triangle is bad is *strictly less* than unity, as we show now.
- This probability is strictly less than $n^6(1 - \frac{k}{n})^s \leq n^6(1 - \frac{1}{r})^{6r \ln n} < n^6 e^{-6 \ln n} = n^6 n^{-6} = 1$
- Therefore, there exists a random sample S of size $s = 6r \log n$ such that the none of the $O(s^2)$ triangles induced by S meet more than $\frac{n}{r}$ lines of L.
- This can be used to design data structures for searching in