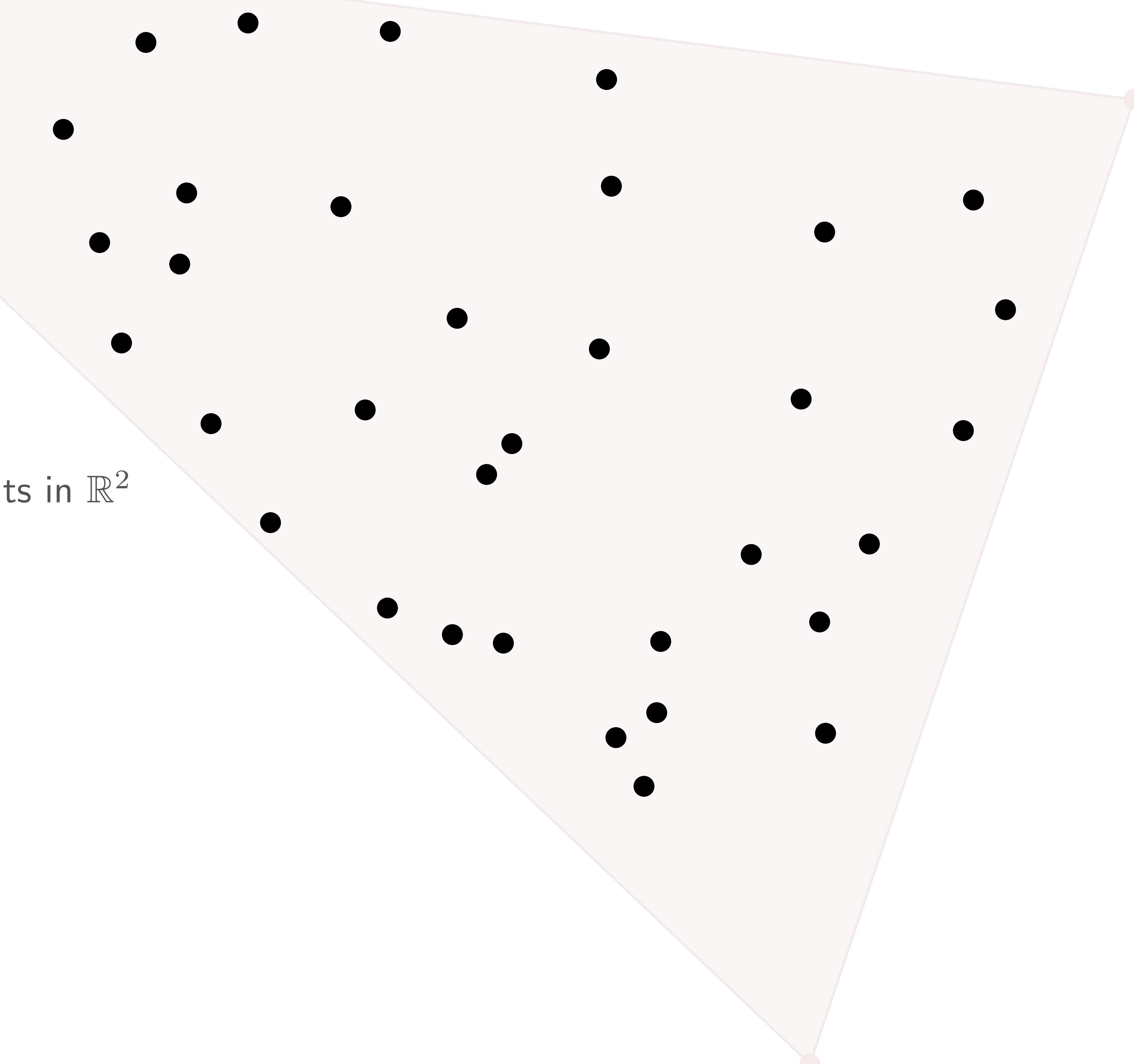
Computing the Expected Area of an Induced Triangle

Vissarion Fisikopoulos — ULB

Frank Staals — MADALGO Center for Massive Data Algorithmics

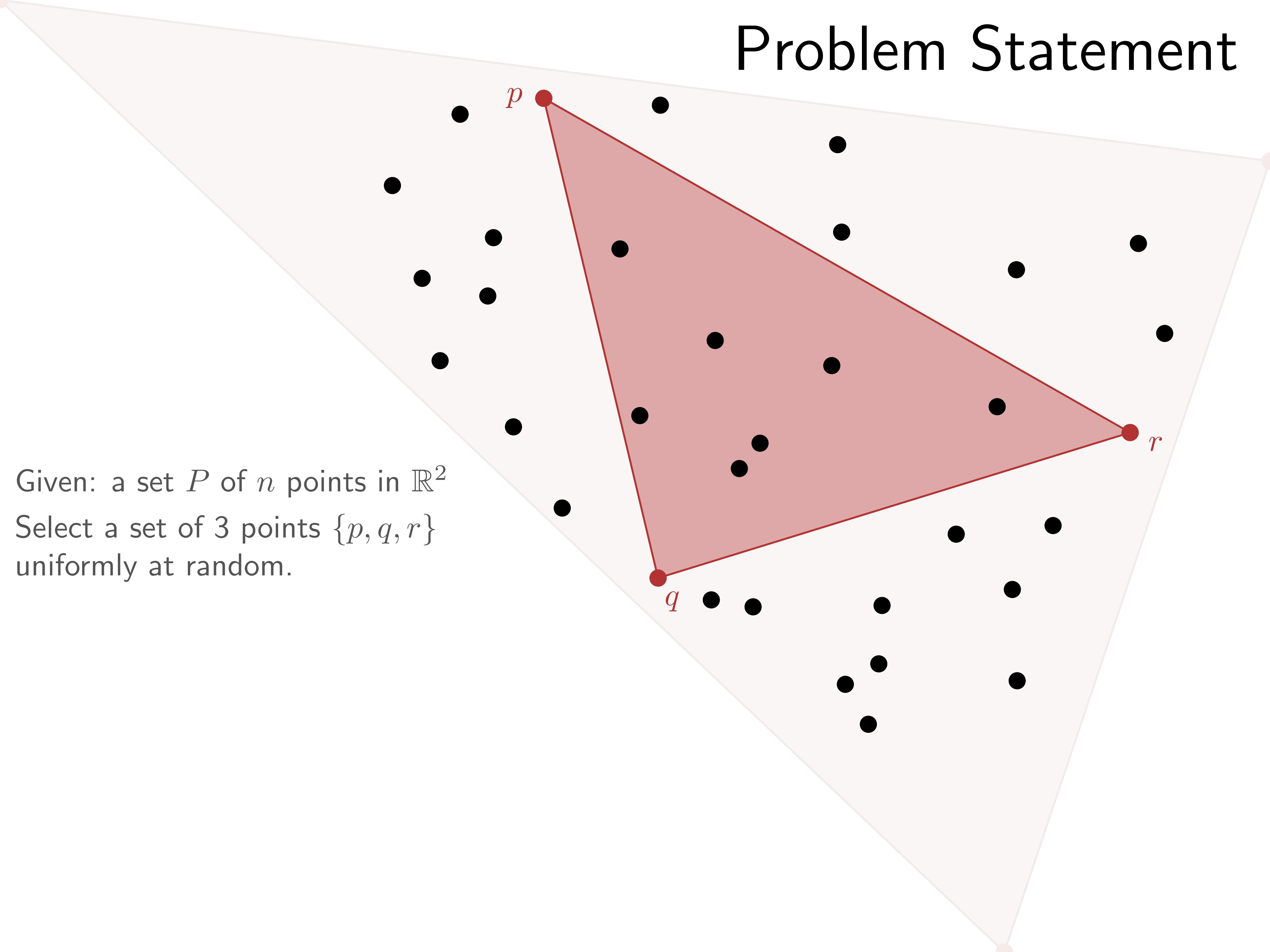
Constantinos Tsirogiannis — MADALGO Center for Massive Data Algorithmics

Problem Statement



Given: a set P of n points in $\mathbb{R}^2$

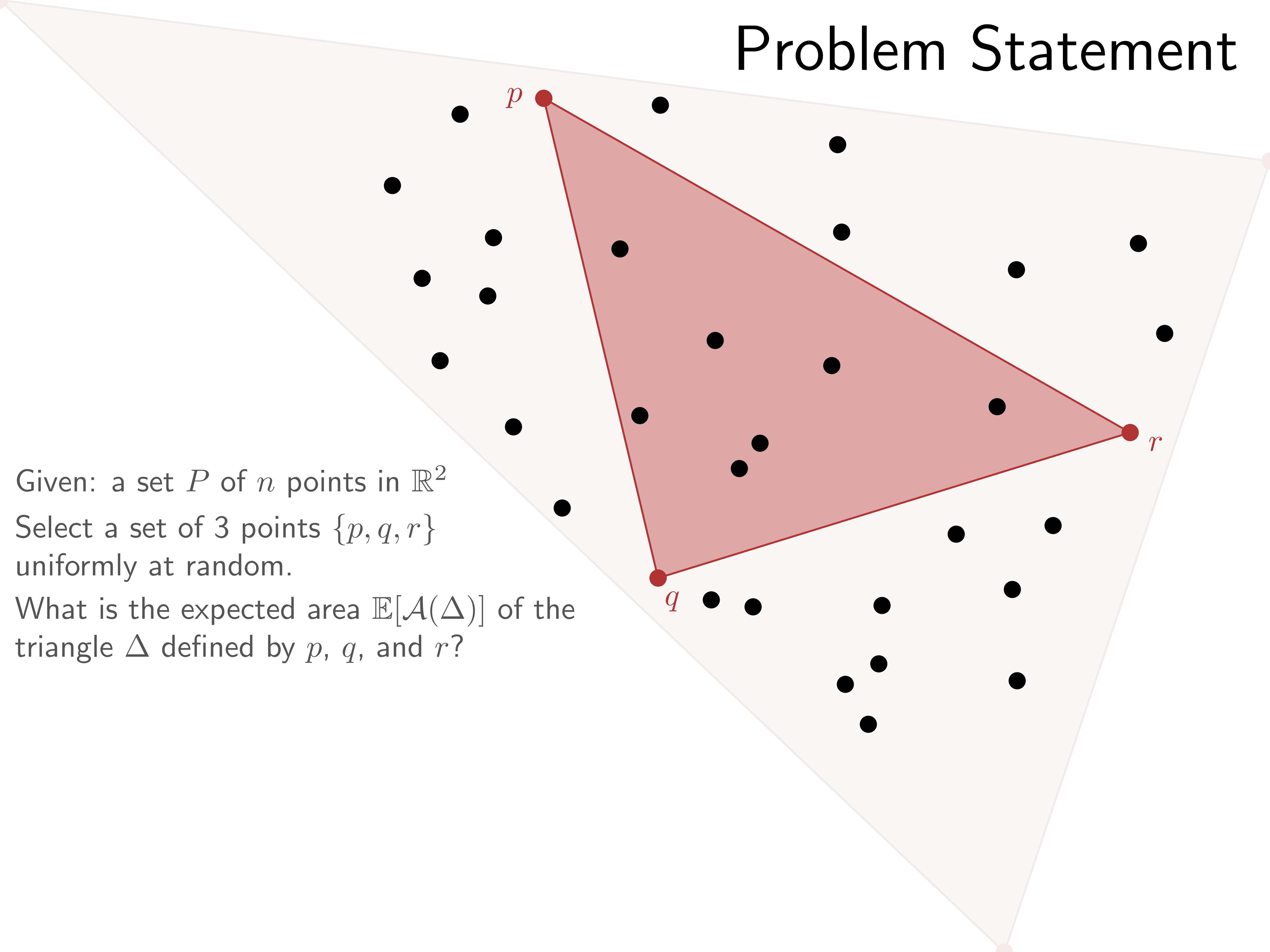
Problem Statement



Given: a set P of n points in $\mathbb{R}^2$

Select a set of 3 points $\{p, q, r\}$ uniformly at random.

Problem Statement

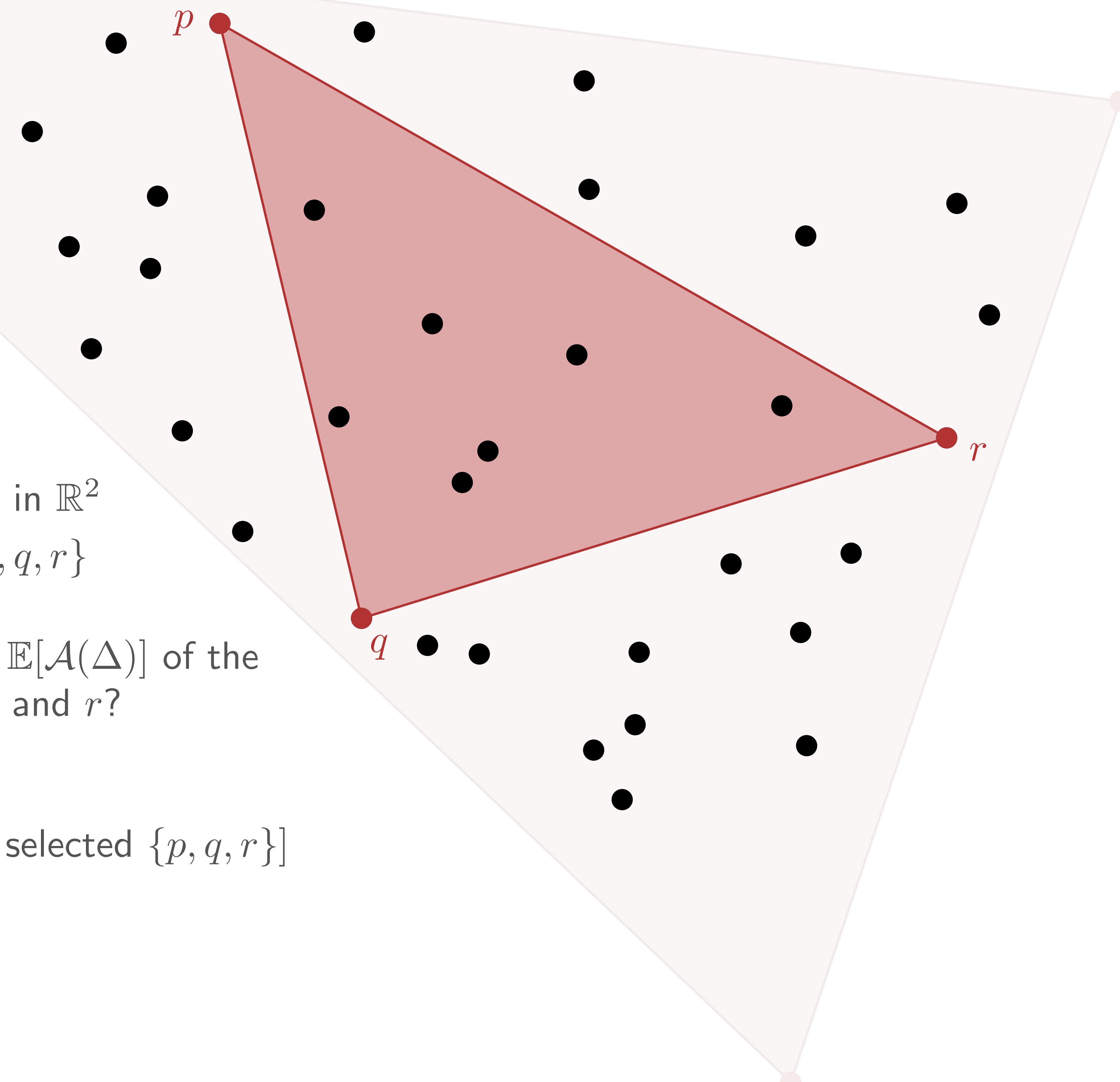


Given: a set P of n points in $\mathbb{R}^2$

Select a set of 3 points $\{p, q, r\}$ uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

Problem Statement



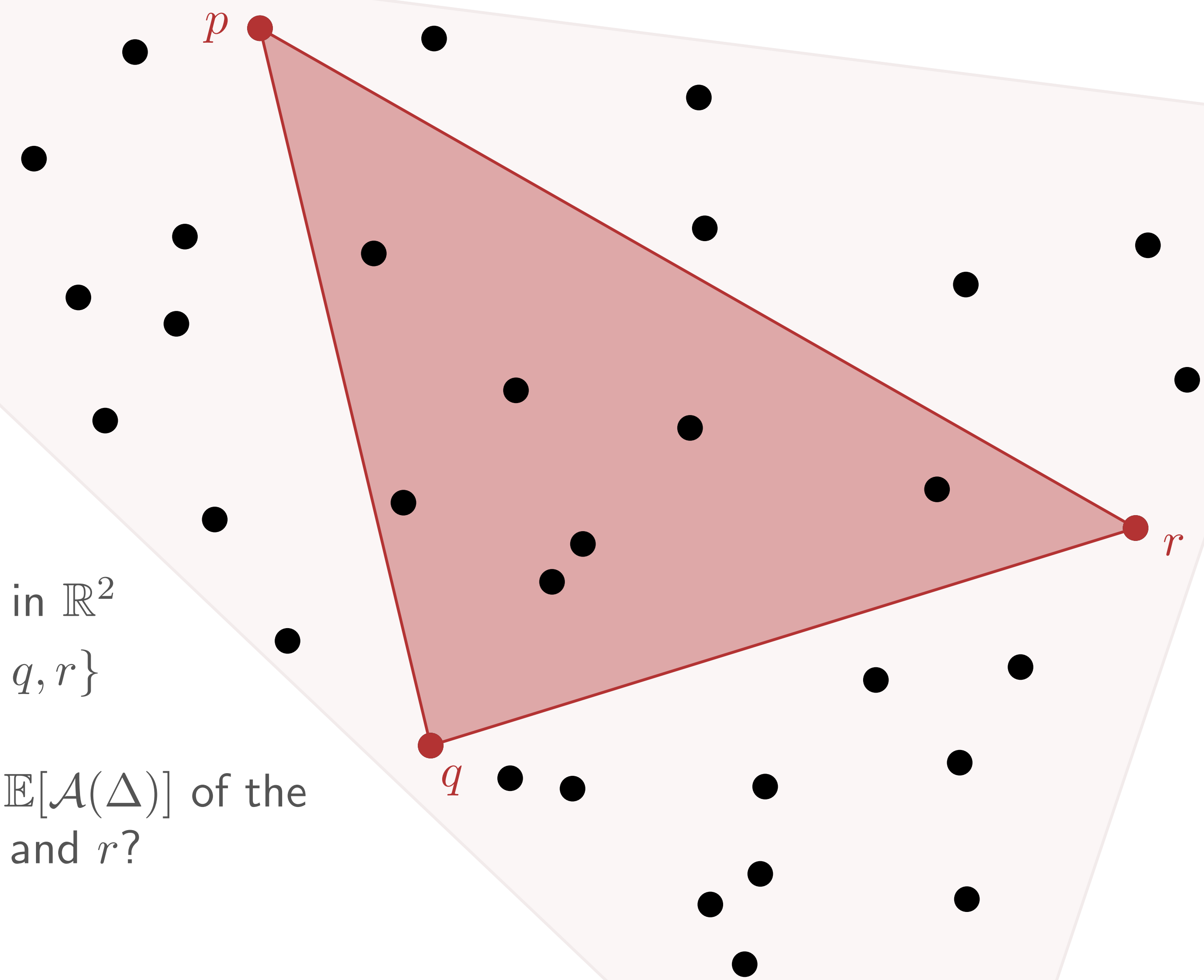
Given: a set P of n points in $\mathbb{R}^2$

Select a set of 3 points $\{p, q, r\}$ uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

$$\sum_{\{p,q,r\} \subseteq P} \mathcal{A}(\Delta)\mathbb{P}[\text{we selected } \{p, q, r\}]$$

Problem Statement



Given: a set P of n points in $\mathbb{R}^2$

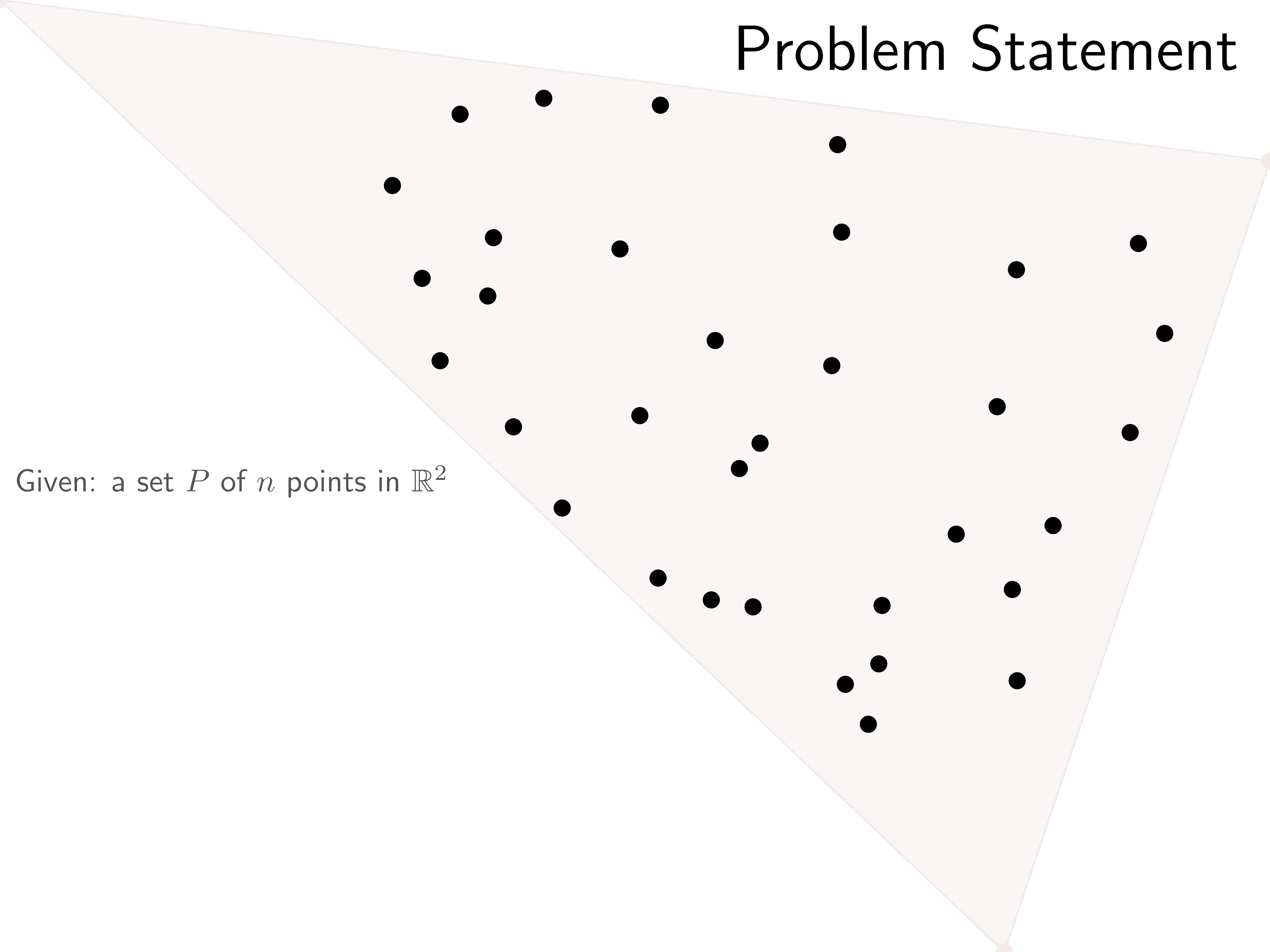
Select a set of 3 points $\{p, q, r\}$ uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

$$\sum_{\{p,q,r\}\subseteq P} \mathcal{A}(\Delta)\mathbb{P}[\text{we selected } \{p, q, r\}]$$

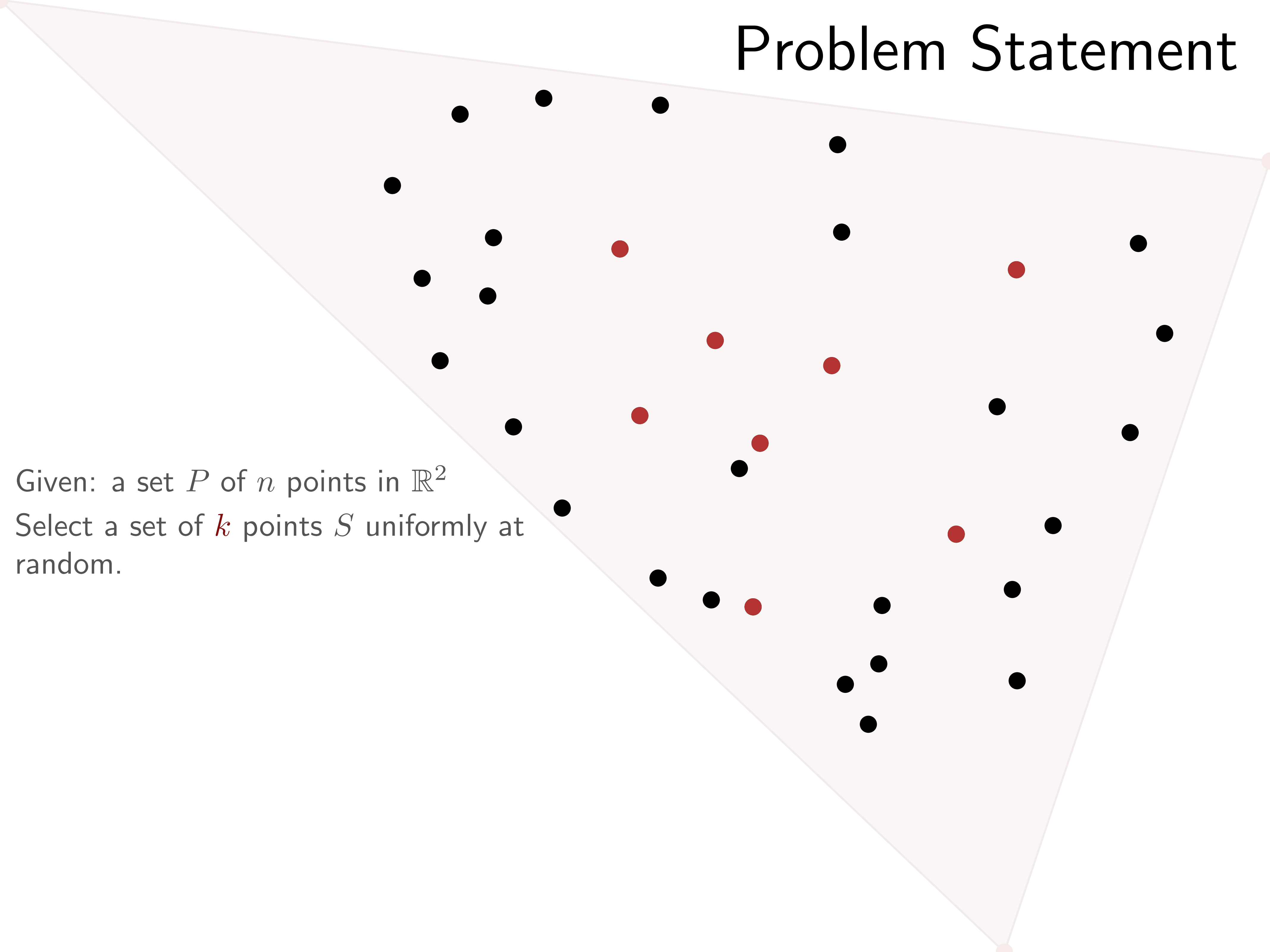
Naive: $O(n^3)$

Problem Statement



Given: a set P of n points in $\mathbb{R}^2$

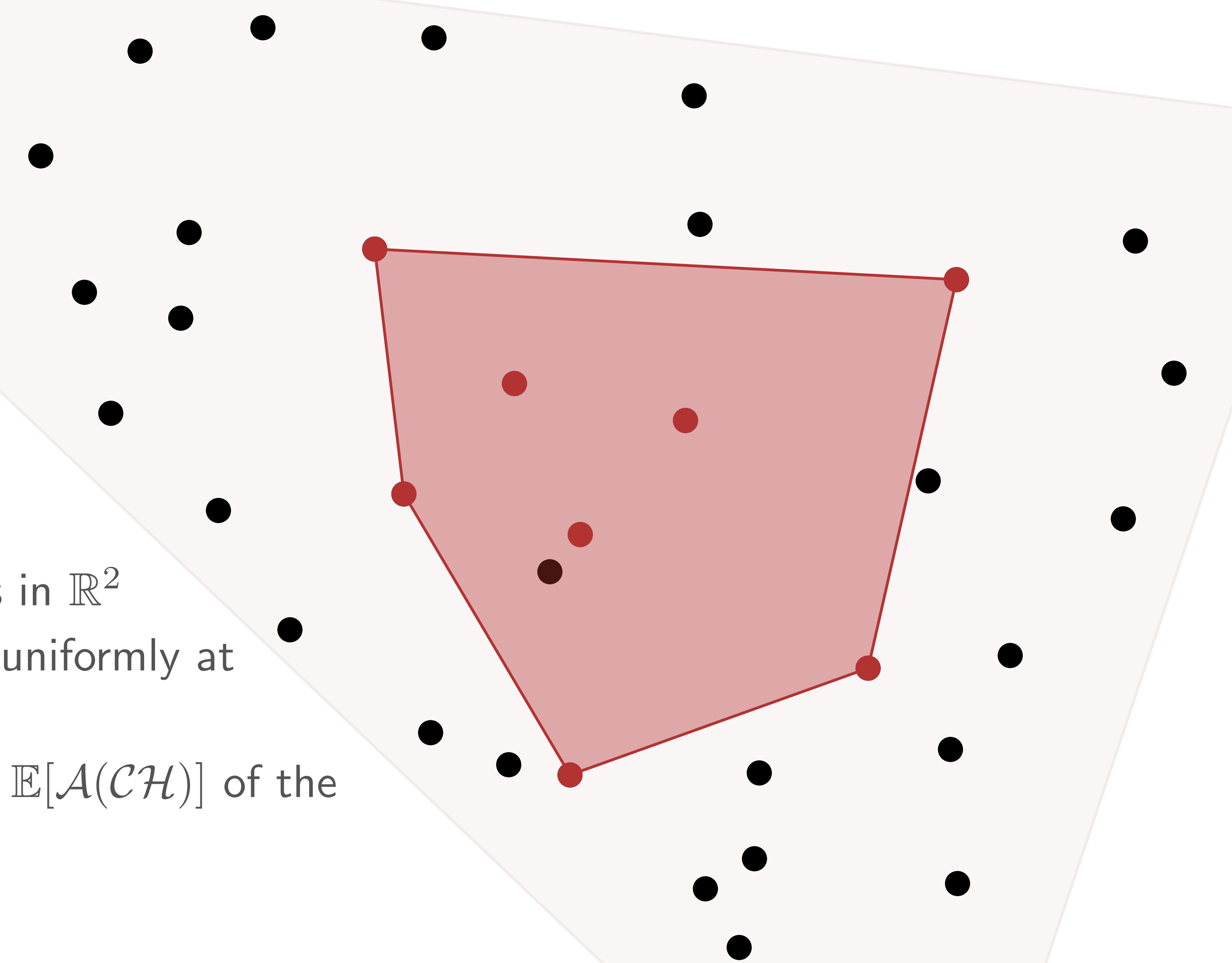
Problem Statement



Given: a set P of n points in $\mathbb{R}^2$

Select a set of k points S uniformly at random.

Problem Statement

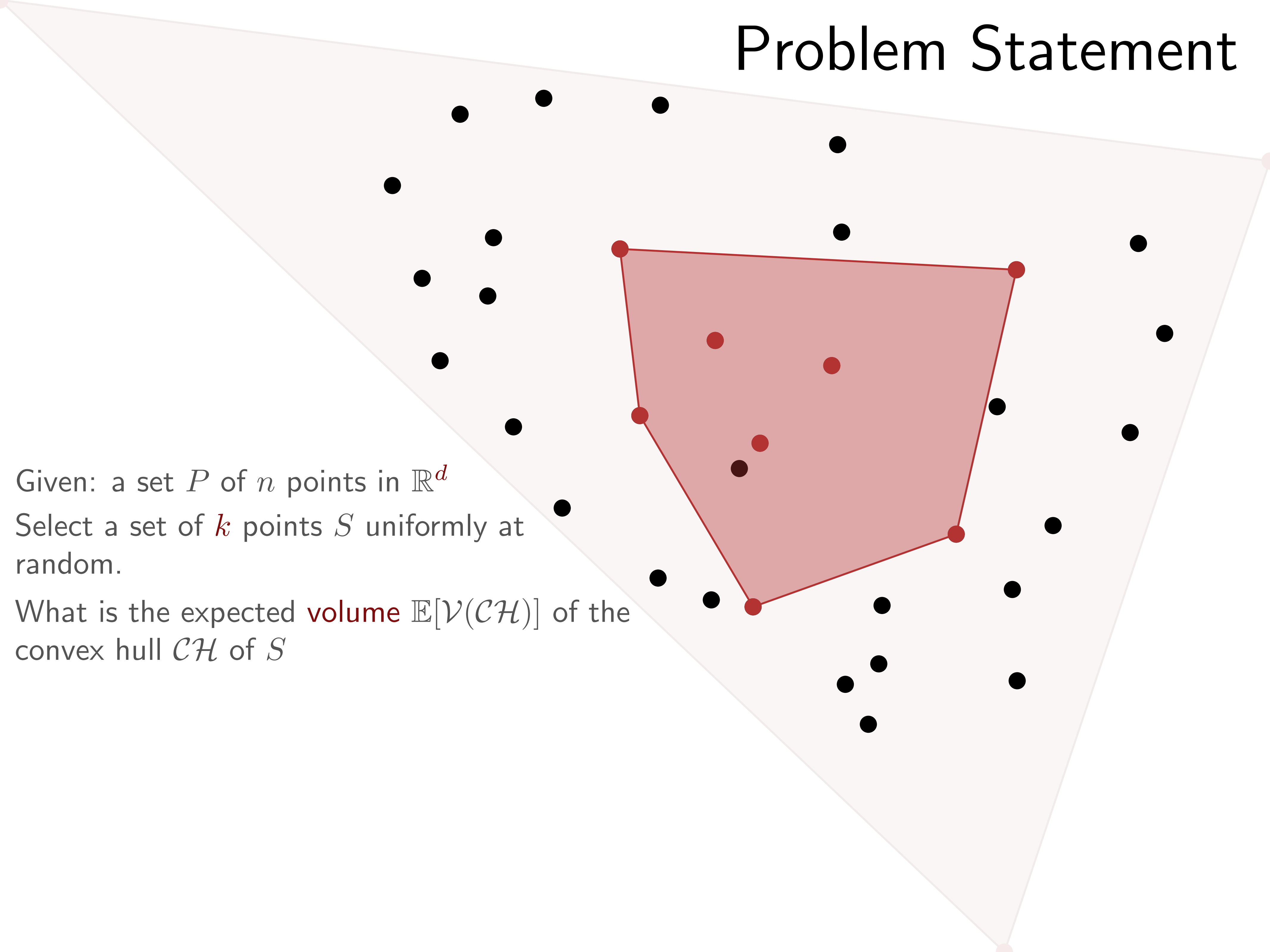


Given: a set P of n points in $\mathbb{R}^2$

Select a set of k points S uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\mathcal{CH})]$ of the convex hull $\mathcal{CH}$ of S

Problem Statement

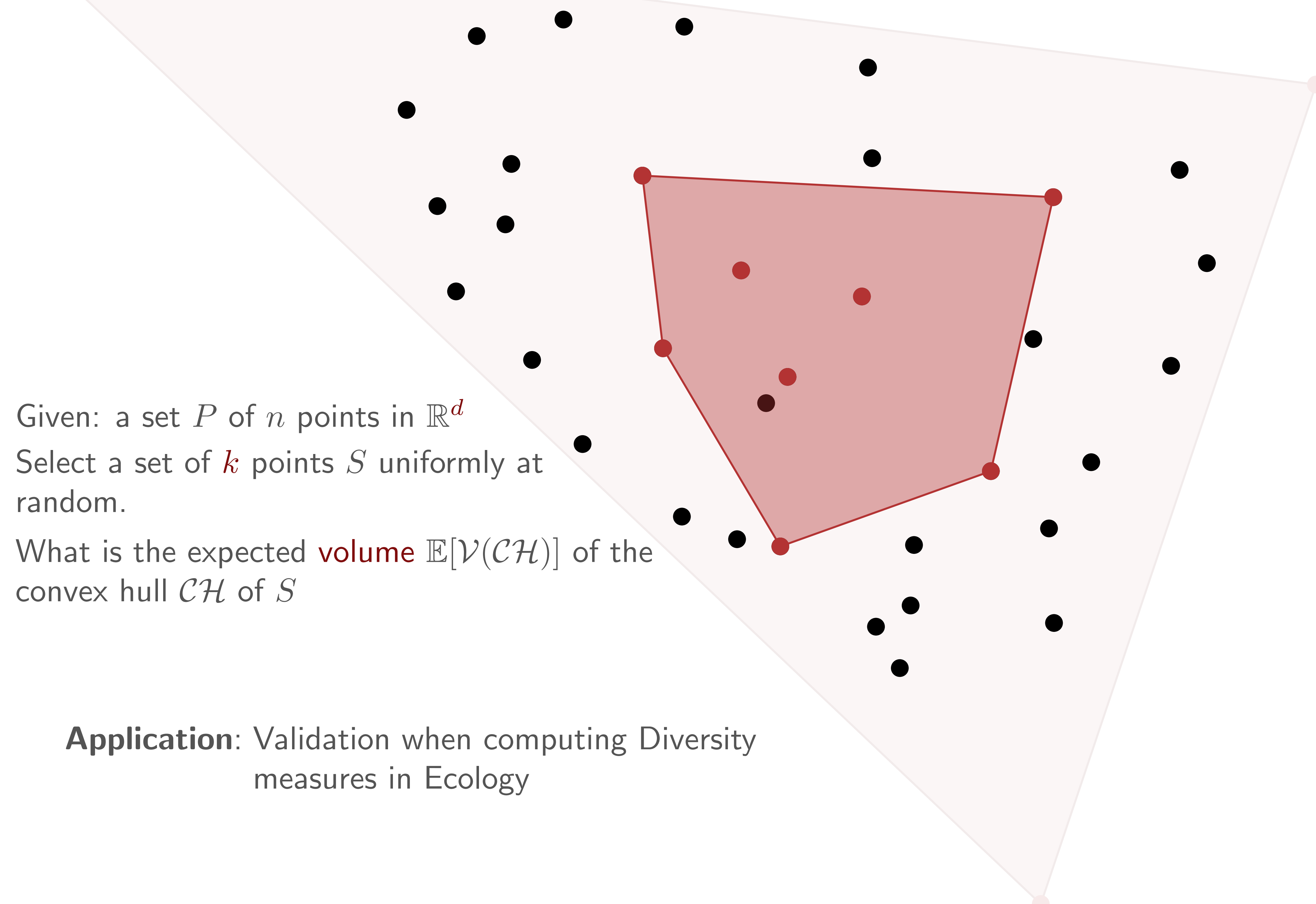


Given: a set P of n points in $\mathbb{R}^d$

Select a set of k points S uniformly at random.

What is the expected volume $\mathbb{E}[\mathcal{V}(\mathcal{CH})]$ of the convex hull $\mathcal{CH}$ of S

Problem Statement



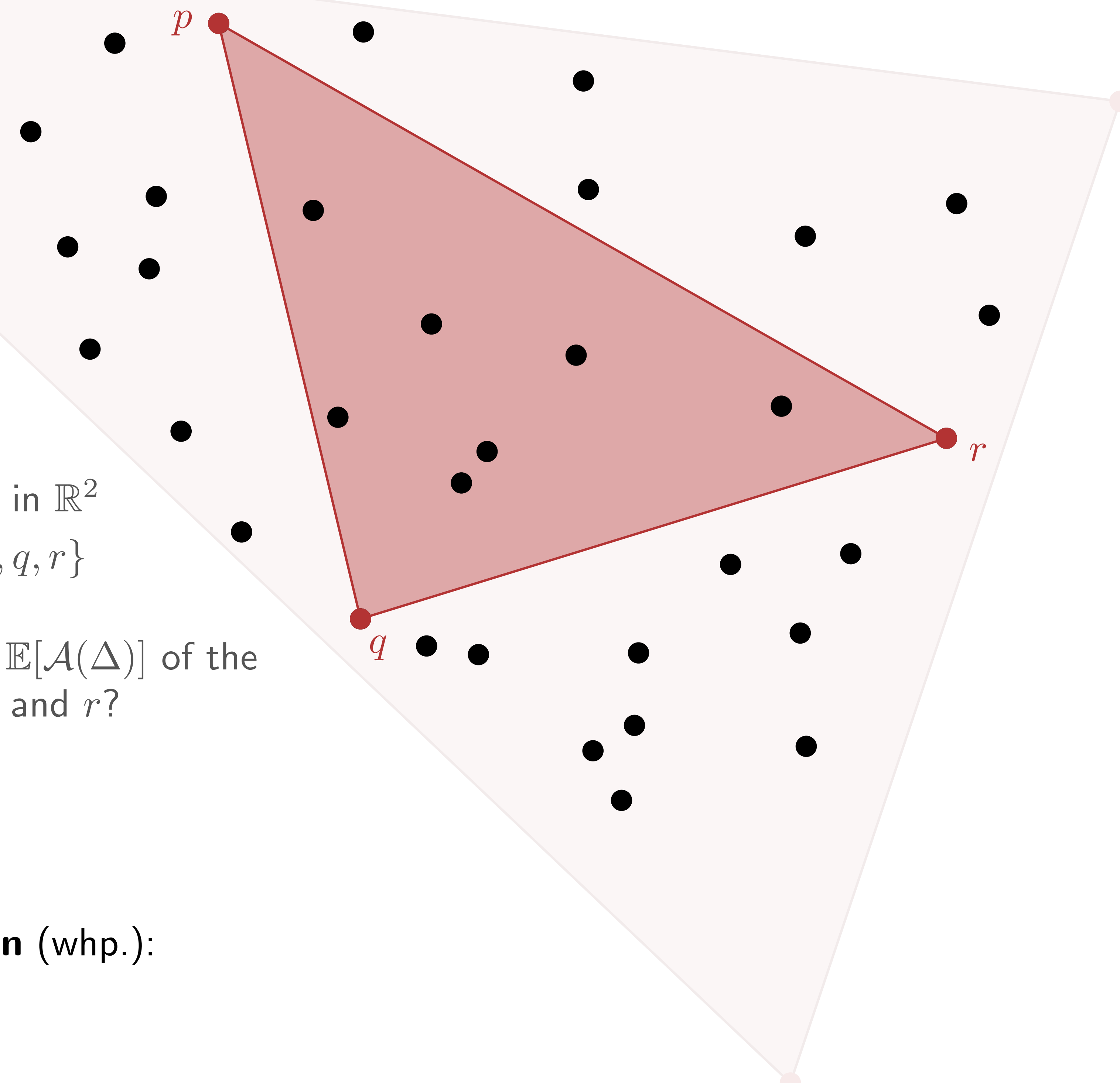
Given: a set P of n points in $\mathbb{R}^d$

Select a set of k points S uniformly at random.

What is the expected volume $\mathbb{E}[\mathcal{V}(\mathcal{CH})]$ of the convex hull $\mathcal{CH}$ of S

Application: Validation when computing Diversity measures in Ecology

Results



Given: a set P of n points in $\mathbb{R}^2$

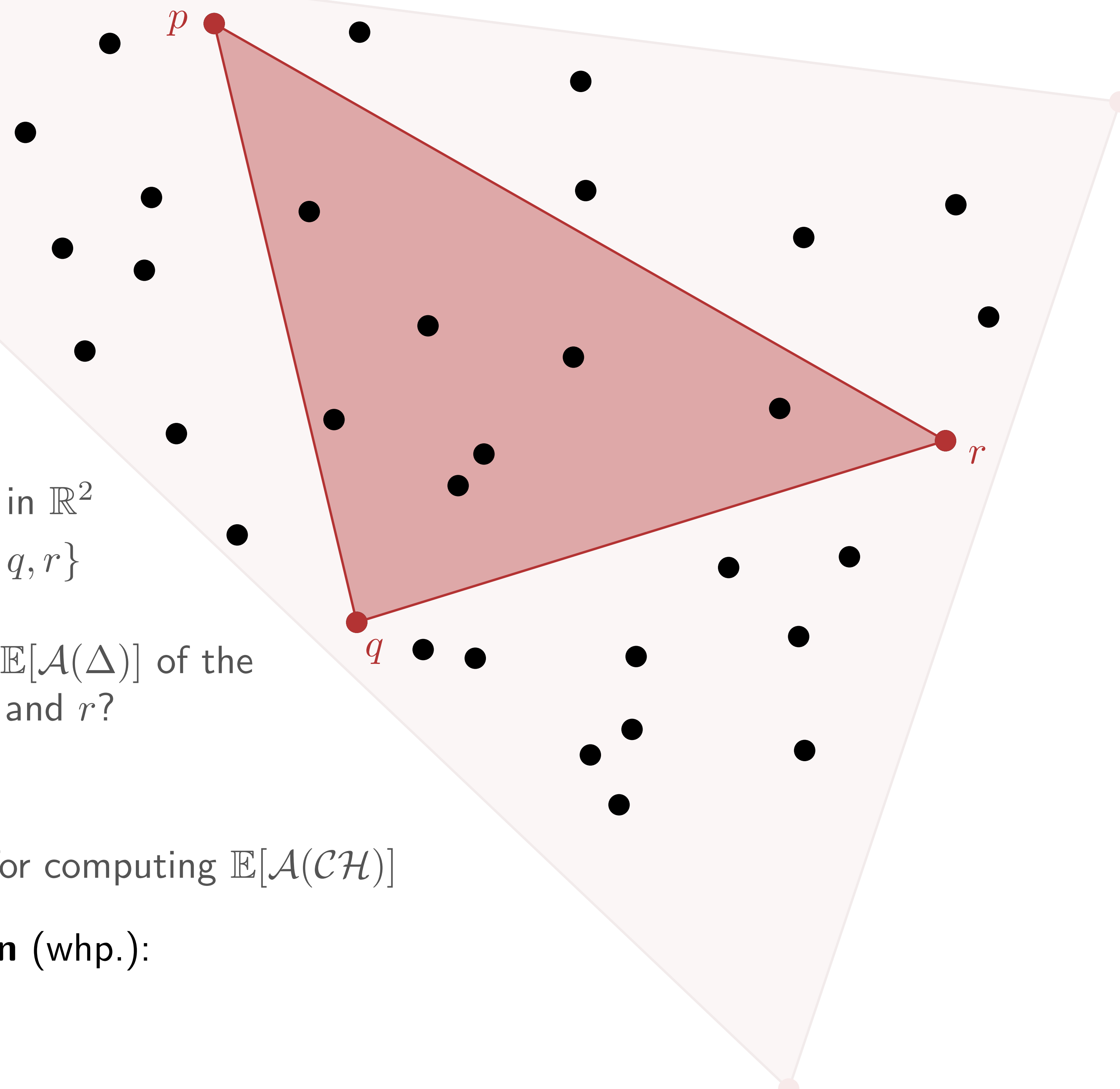
Select a set of 3 points $\{p, q, r\}$ uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

Exact: $O(n^2 \log n)$

$(1 \pm \varepsilon)$-**Approximation** (whp.):

Results



Given: a set P of n points in $\mathbb{R}^2$

Select a set of 3 points $\{p, q, r\}$ uniformly at random.

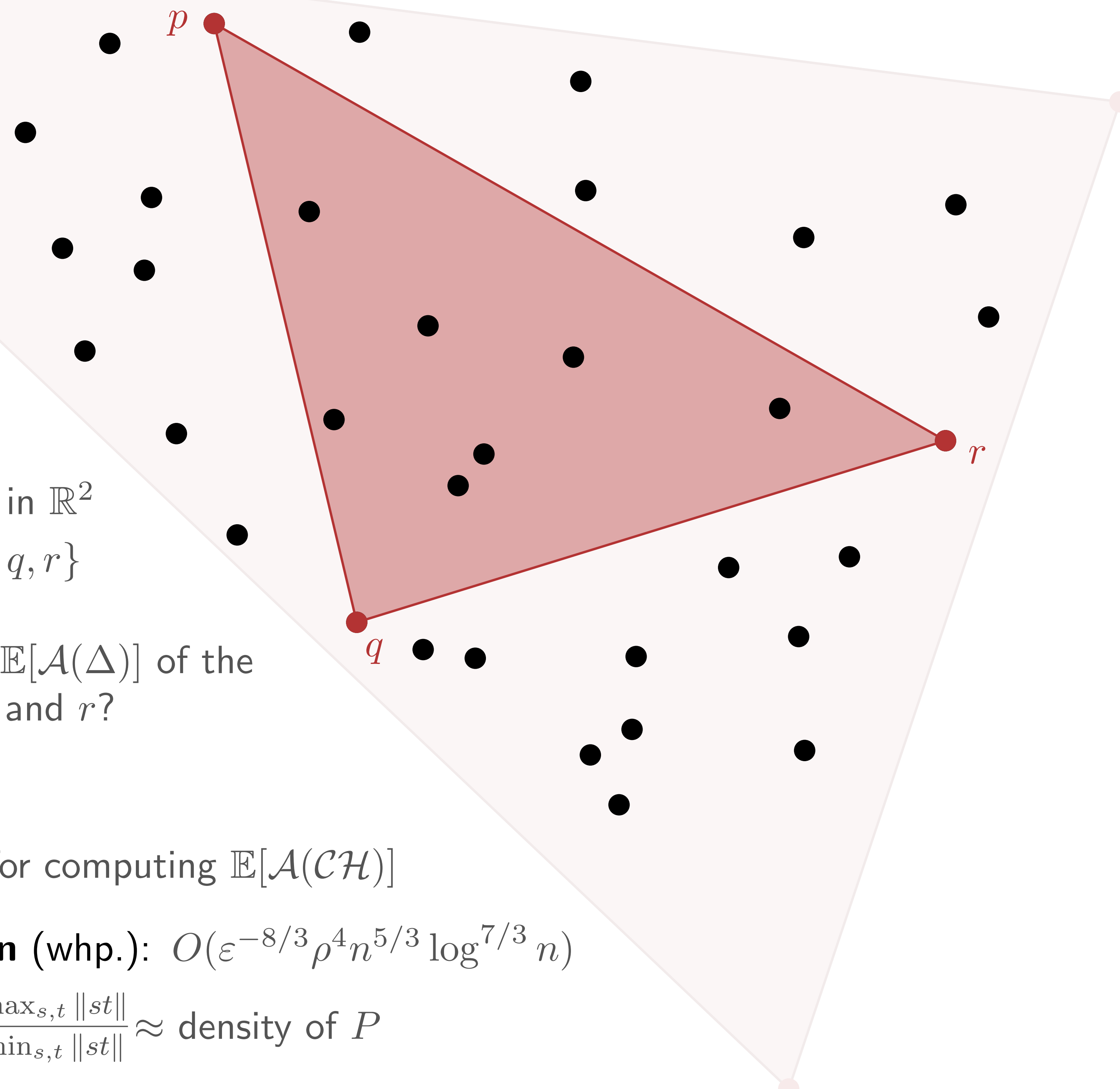
What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

Exact: $O(n^2 \log n)$

Also works for computing $\mathbb{E}[\mathcal{A}(\mathcal{CH})]$

$(1 \pm \varepsilon)$-**Approximation** (whp.):

Results



Given: a set P of n points in $\mathbb{R}^2$

Select a set of 3 points $\{p, q, r\}$ uniformly at random.

What is the expected area $\mathbb{E}[\mathcal{A}(\Delta)]$ of the triangle Δ defined by p, q, and r?

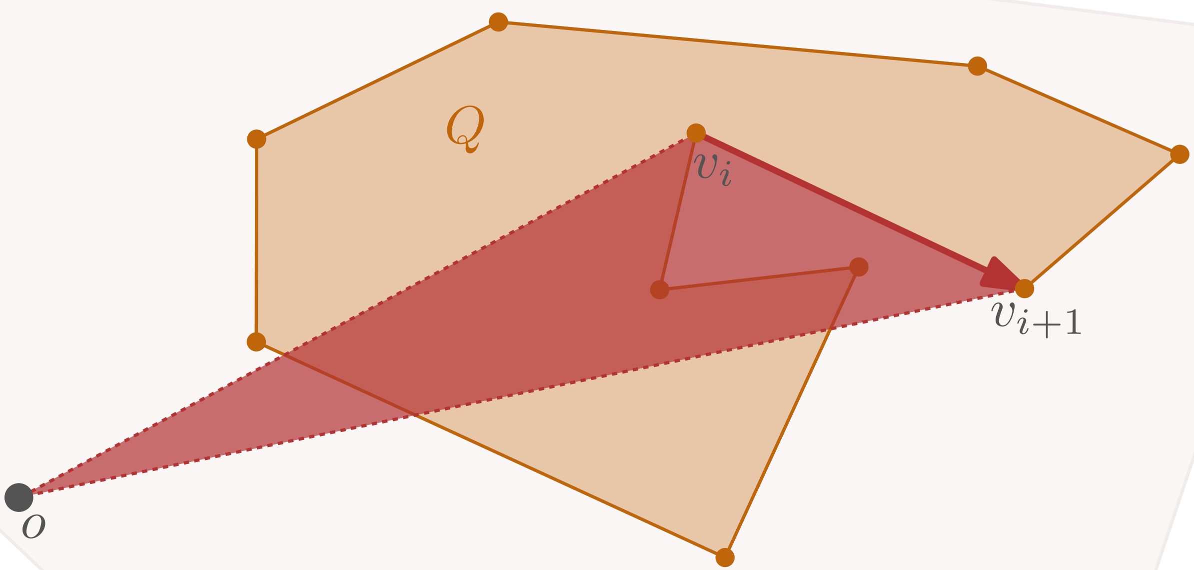
Exact: $O(n^2 \log n)$

Also works for computing $\mathbb{E}[\mathcal{A}(\mathcal{CH})]$

$(1 \pm \varepsilon)$-**Approximation** (whp.): $O(\varepsilon^{-8/3} \rho^4 n^{5/3} \log^{7/3} n)$

$\rho = \frac{\max_{s,t} \|st\|}{\min_{s,t} \|st\|} \approx$ density of P

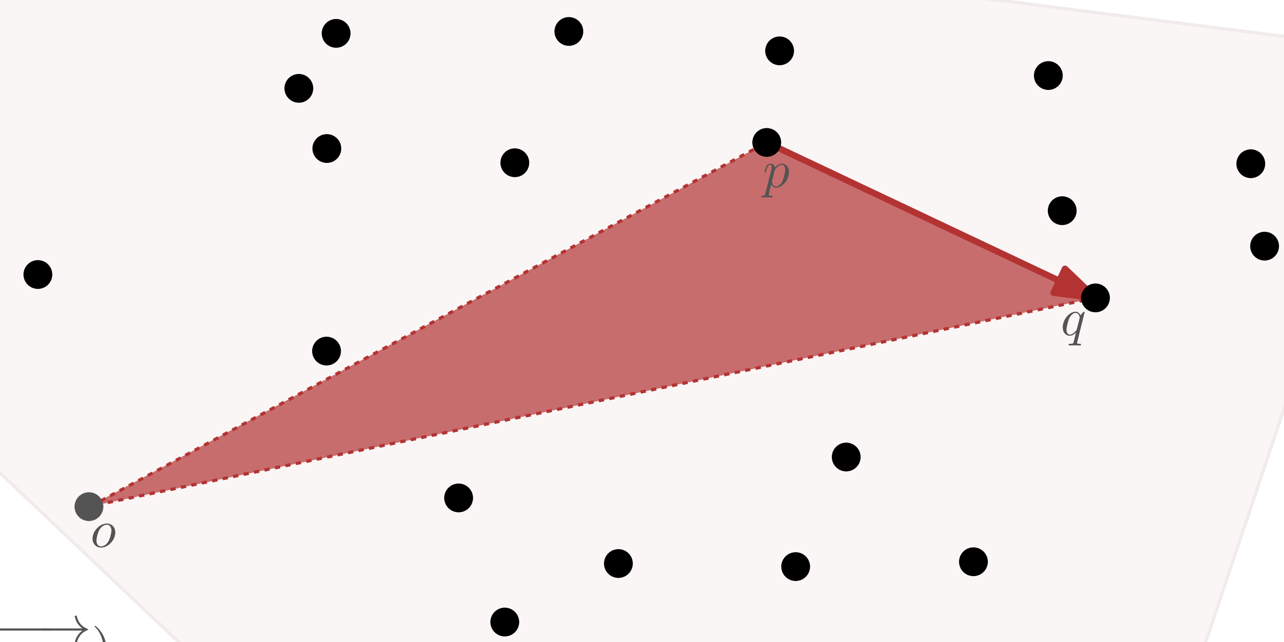
Exact Algorithm



Shoelace formula

$$\mathcal{A}(Q) = \frac{1}{2}\sum_{i=0}^{n} \mathcal{A}'(\overrightarrow{v_i v_{i+1 \mod n}})$$

Exact Algorithm

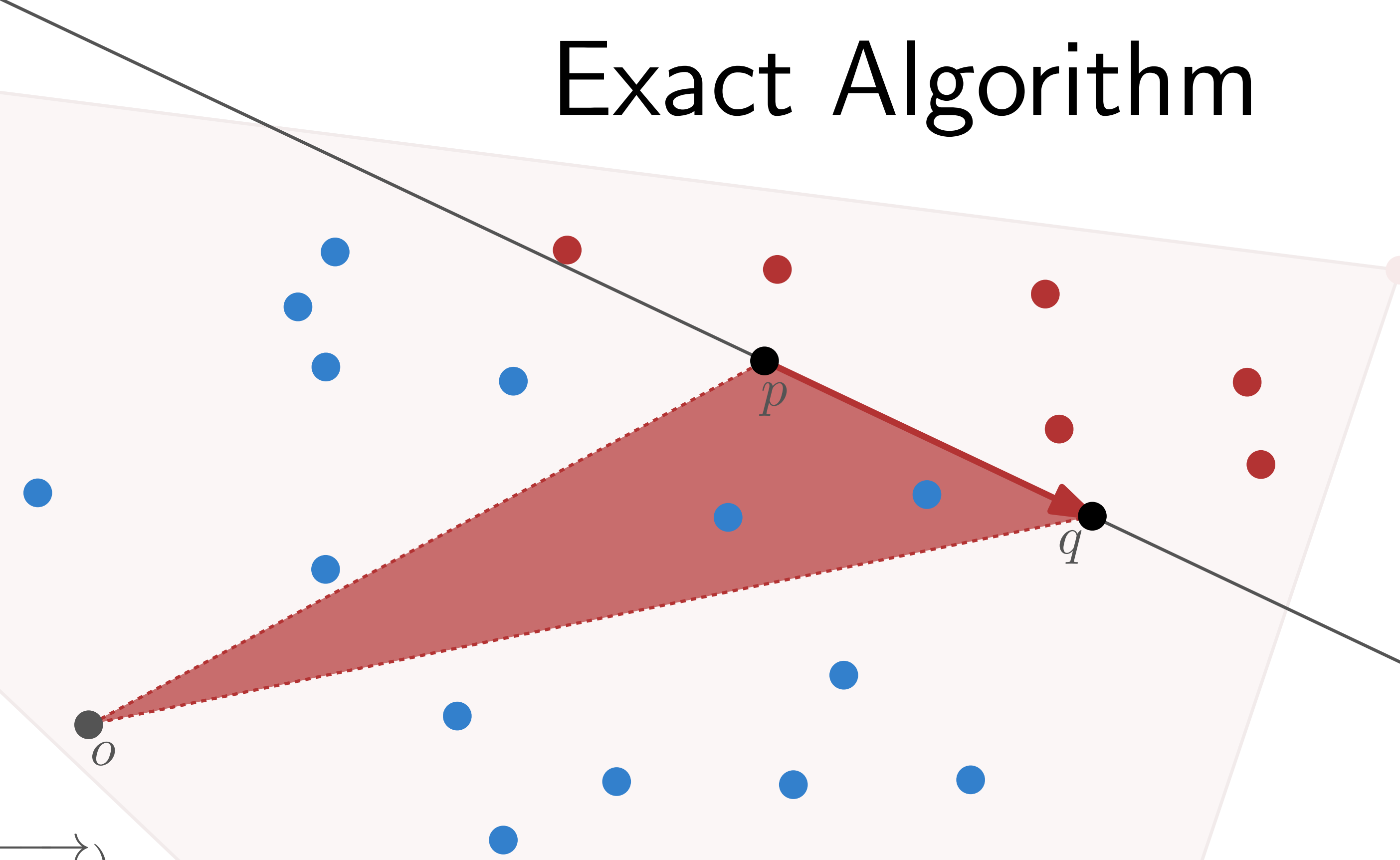


Shoelace formula

$$\mathcal{A}(Q) = \frac{1}{2}\sum_{i=0}^{n} \mathcal{A}'(\overrightarrow{v_i v_{i+1 \mod n}})$$

$$\mathbb{E}[\mathcal{A}(\Delta)] = \sum_{p,q} \mathcal{A}'(\overrightarrow{pq})\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta]$$

Exact Algorithm



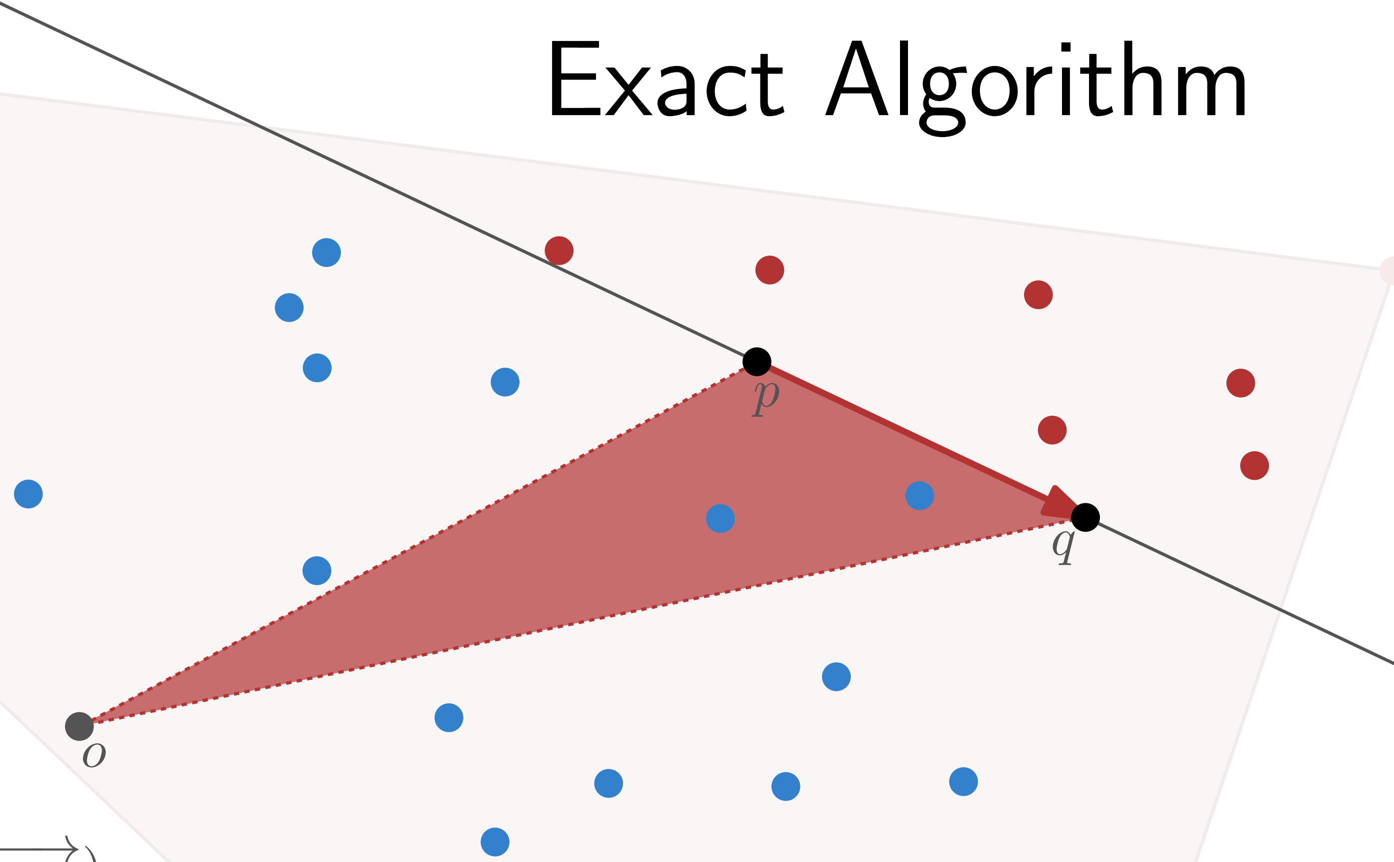
Shoelace formula

$$\mathcal{A}(Q) = \frac{1}{2}\sum_{i=0}^{n} \mathcal{A}'(\overrightarrow{v_i v_{i+1 \bmod n}})$$

$$\mathbb{E}[\mathcal{A}(\Delta)] = \sum_{p,q} \mathcal{A}'(\overrightarrow{pq})\,\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta]$$

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta] = \frac{\binom{n_q}{3-2}}{\binom{n}{3}} = \frac{n_q}{\binom{n}{3}}$$

Exact Algorithm



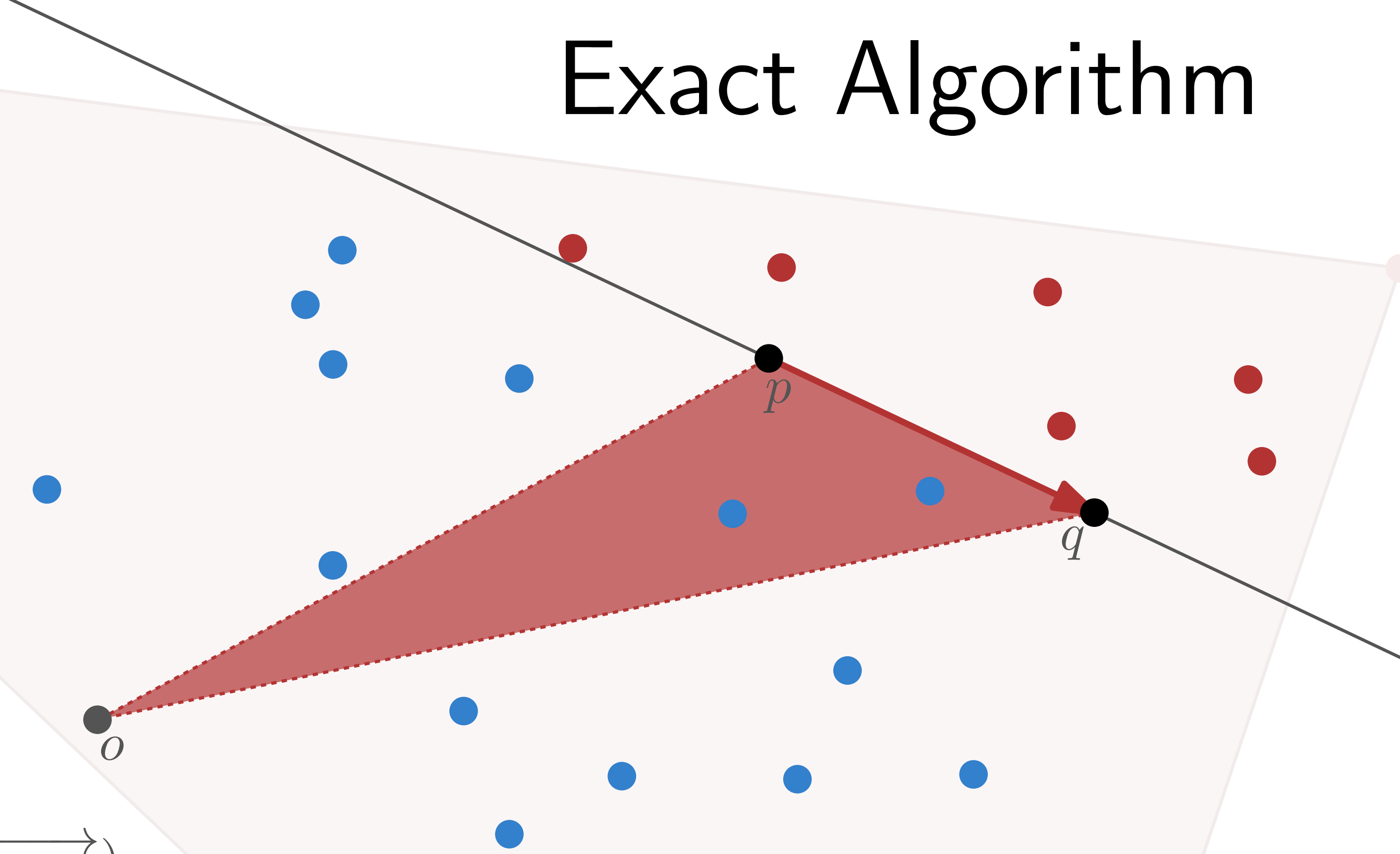
Shoelace formula

$$\mathcal{A}(Q) = \frac{1}{2}\sum_{i=0}^{n} \mathcal{A}'(\overrightarrow{v_i v_{i+1 \mod n}})$$

$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq}) n_{pq}$$

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta] = \frac{\binom{n_q}{3-2}}{\binom{n}{3}} = \frac{n_q}{\binom{n}{3}}$$

Exact Algorithm



Shoelace formula

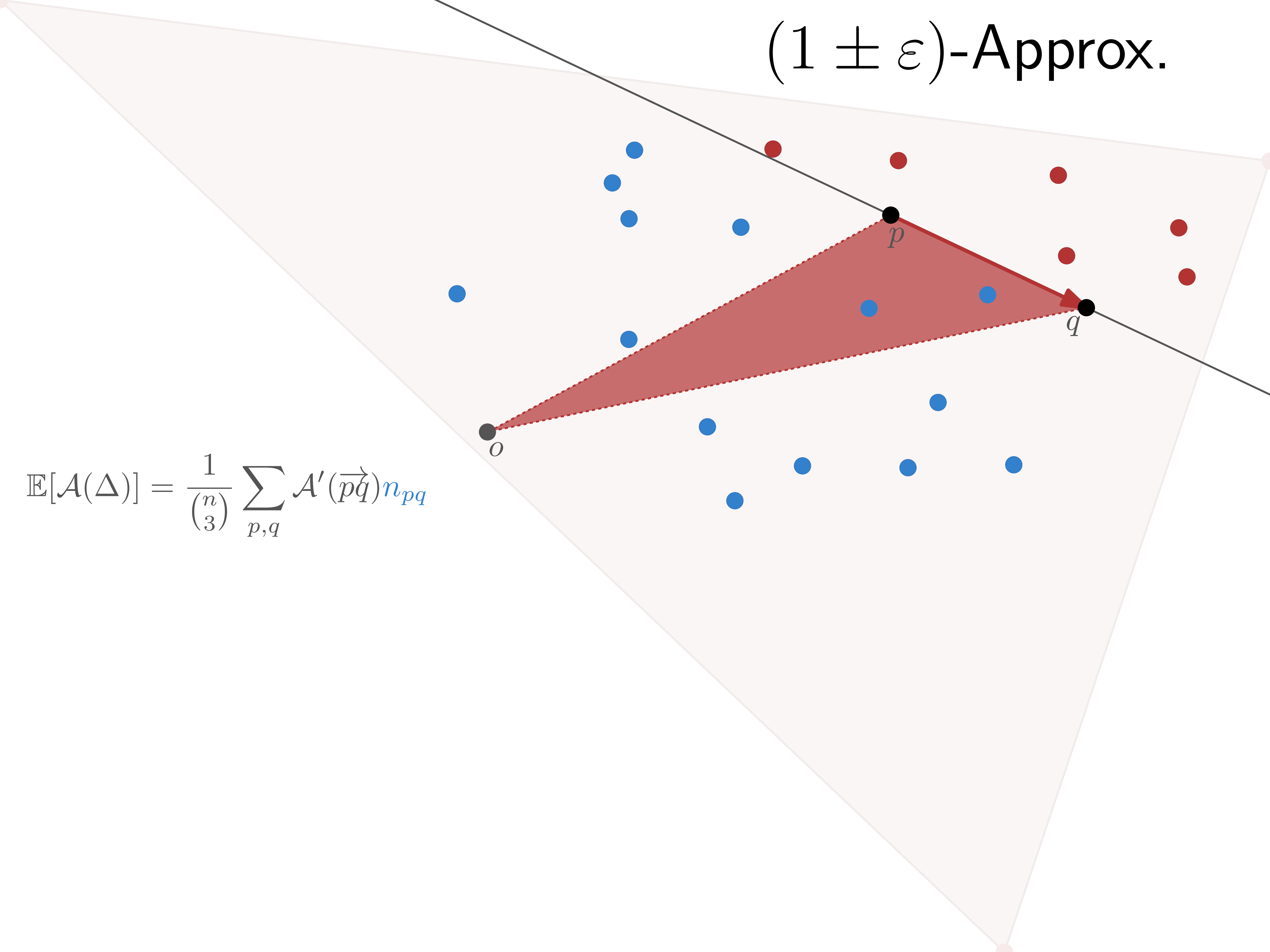
$$\mathcal{A}(Q) = \frac{1}{2}\sum_{i=0}^{n} \mathcal{A}'(\overrightarrow{v_i v_{i+1 \mod n}})$$

$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq})\, n_{pq}$$

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta] = \frac{\binom{n_q}{3-2}}{\binom{n}{3}} = \frac{n_q}{\binom{n}{3}}$$

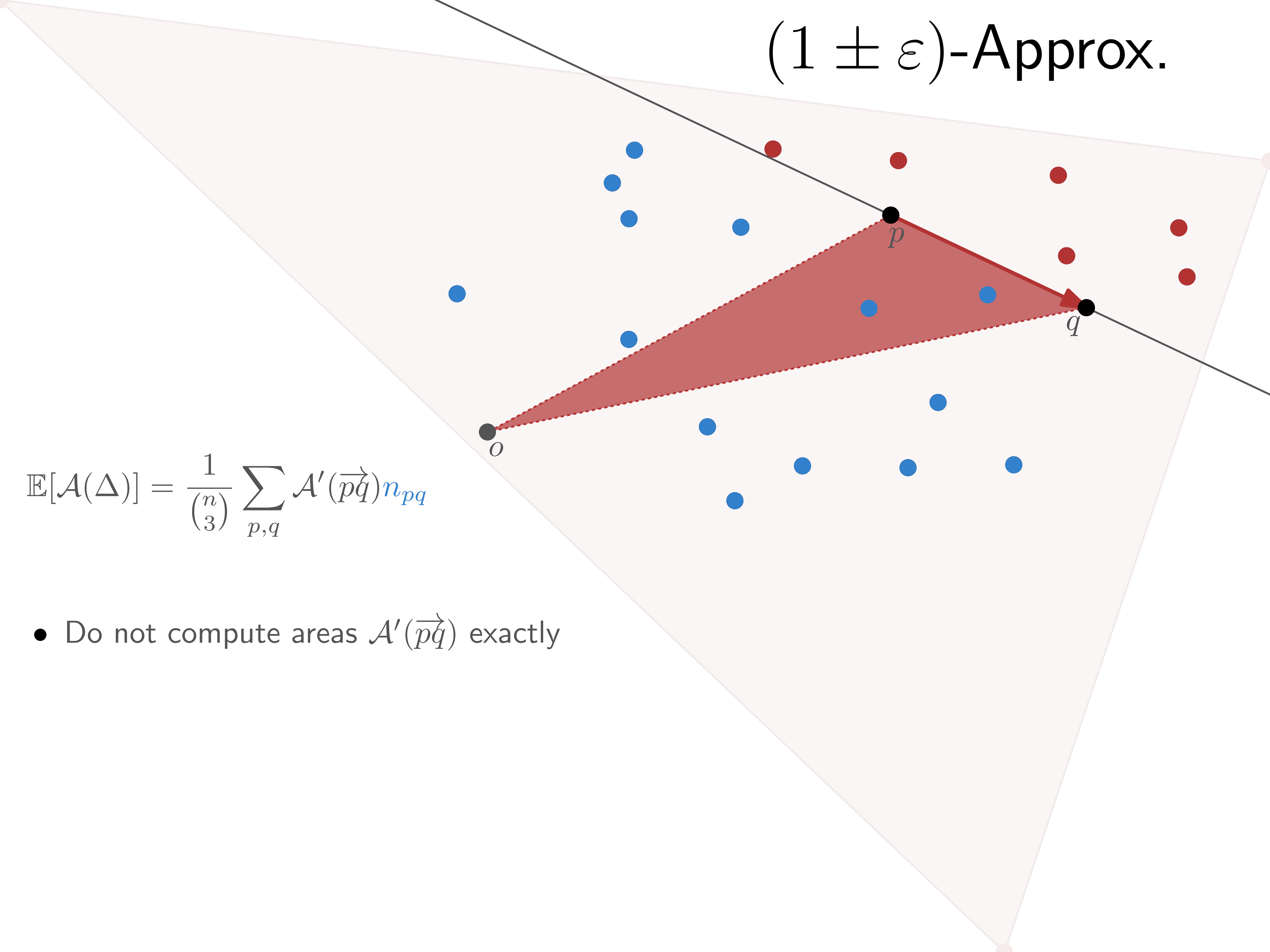
Theorem.
We can compute $\mathbb{E}[\mathcal{A}(\Delta)]$ in $O(n^2 \log n)$ time.

$(1 \pm \varepsilon)$-Approx.



$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq})\, n_{pq}$$

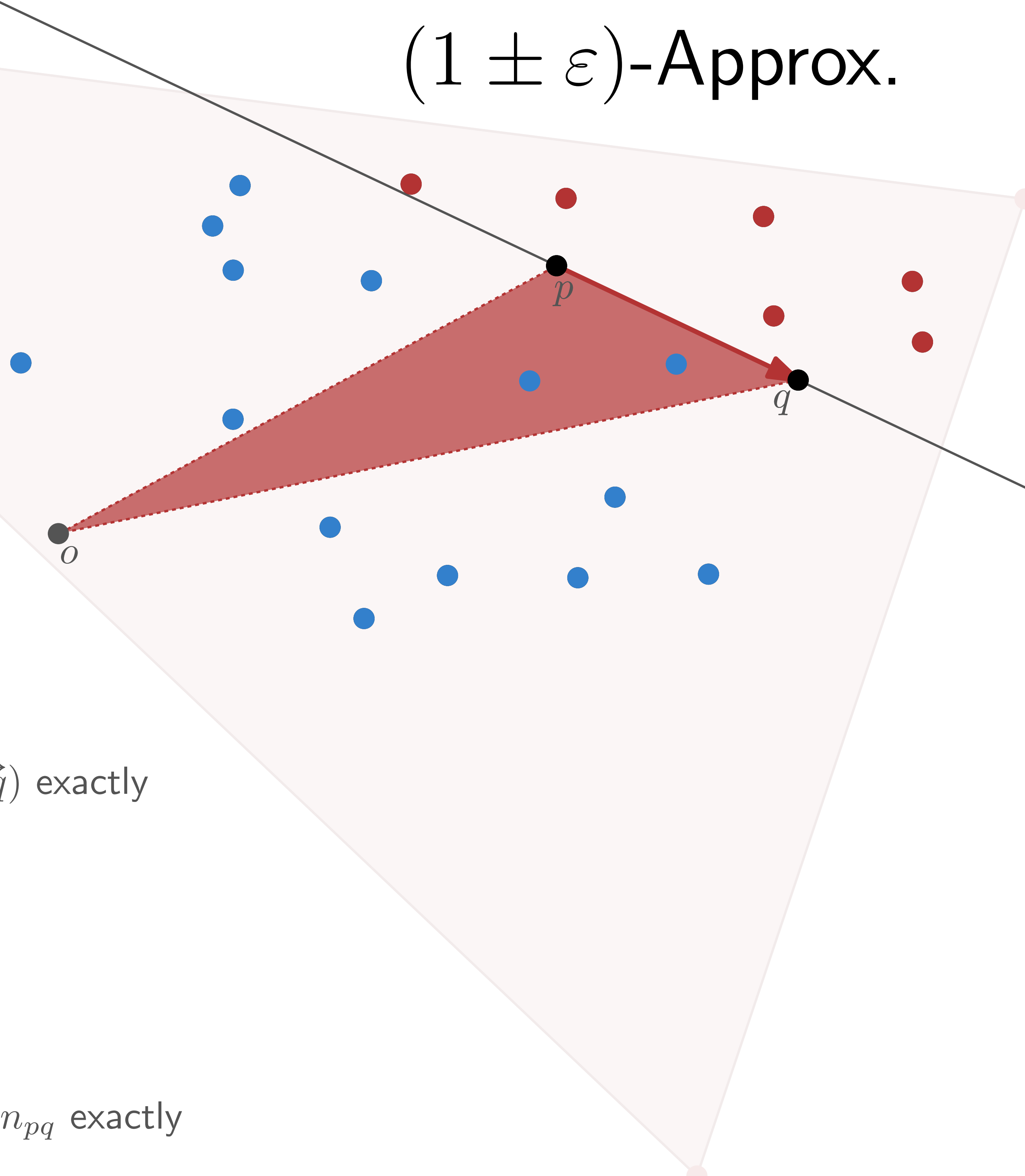
$(1 \pm \varepsilon)$-Approx.



$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq}) n_{pq}$$

- Do not compute areas $\mathcal{A}'(\overrightarrow{pq})$ exactly

$(1 \pm \varepsilon)$-Approx.

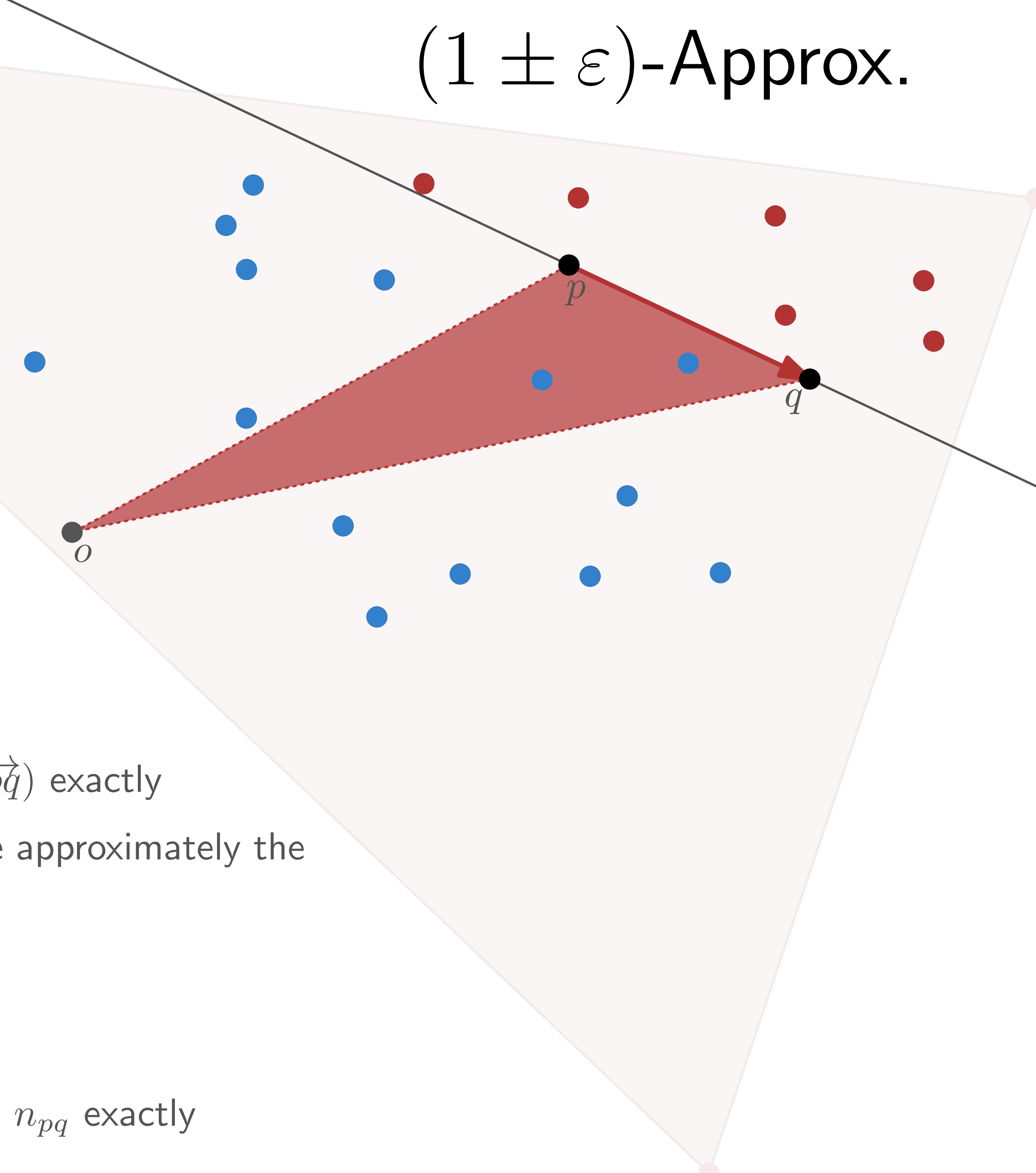


$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq}) n_{pq}$$

- Do not compute areas $\mathcal{A}'(\overrightarrow{pq})$ exactly

- Do not compute the counts n_{pq} exactly

$(1 \pm \varepsilon)$-Approx.



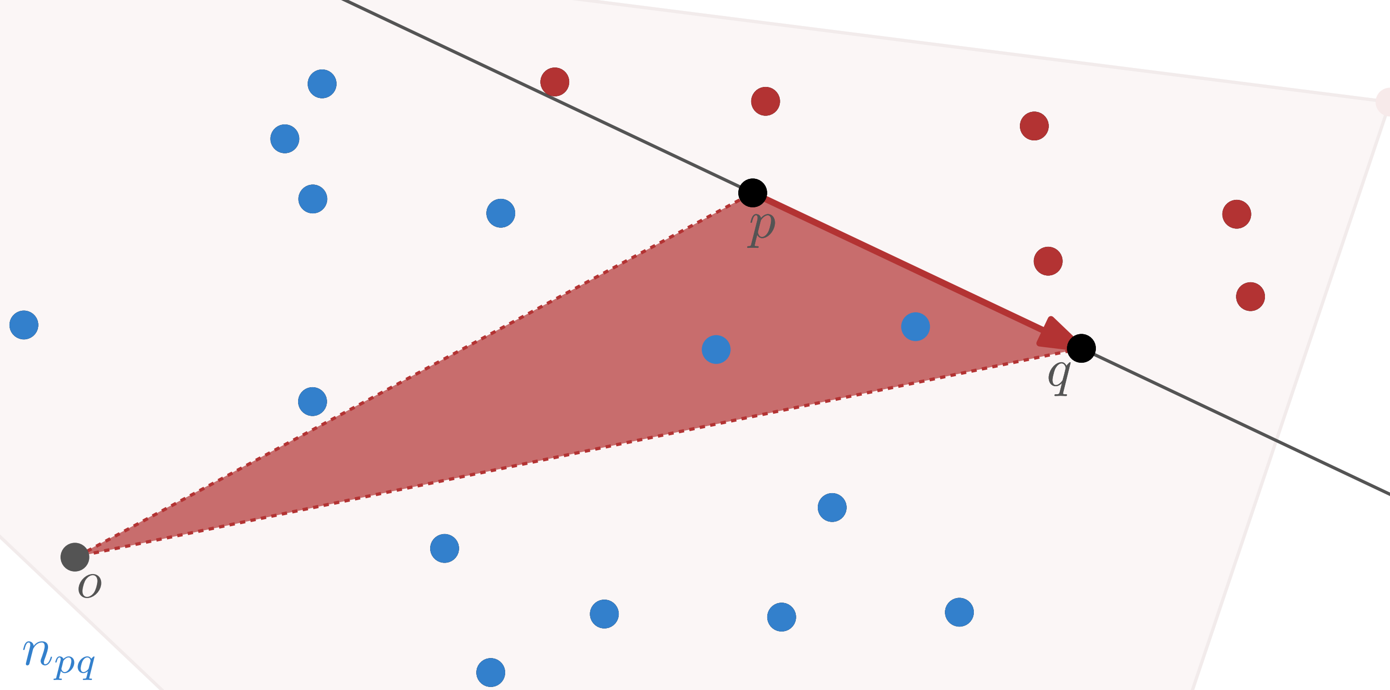
$$\mathbb{E}[\mathcal{A}(\Delta)] = \frac{1}{\binom{n}{3}} \sum_{p,q} \mathcal{A}'(\overrightarrow{pq}) n_{pq}$$

- Do not compute areas $\mathcal{A}'(\overrightarrow{pq})$ exactly

  Group pairs p, q that have approximately the same $\mathcal{A}'(\overrightarrow{pq})$

- Do not compute the counts n_{pq} exactly

$(1 \pm \varepsilon)$-Approx.



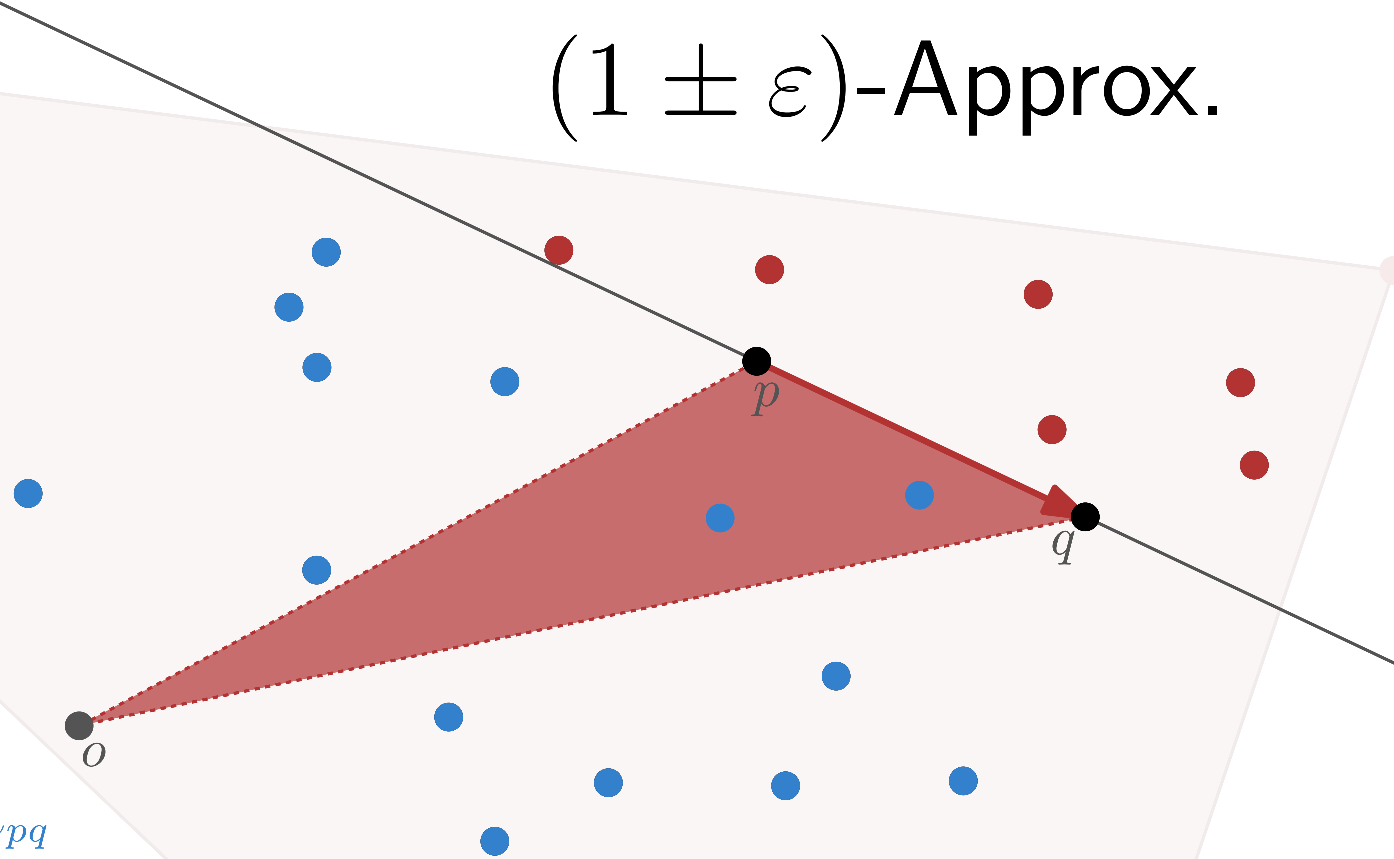
$$\mathbb{E}[\mathcal{A}(\Delta)] \approx \frac{1}{\binom{n}{3}} \sum_{G} \mathcal{A}'(G) \sum_{p,q \in G} n_{pq}$$

- Do not compute areas $\mathcal{A}'(\overrightarrow{pq})$ exactly

  Group pairs p, q that have approximately the same $\mathcal{A}'(\overrightarrow{pq})$

- Do not compute the counts n_{pq} exactly

$(1 \pm \varepsilon)$-Approx.



$$\mathbb{E}[\mathcal{A}(\Delta)] \approx \frac{1}{\binom{n}{3}} \sum_{(p,Q)} A_{pQ} \sum_{q \in Q} n_{pq}$$

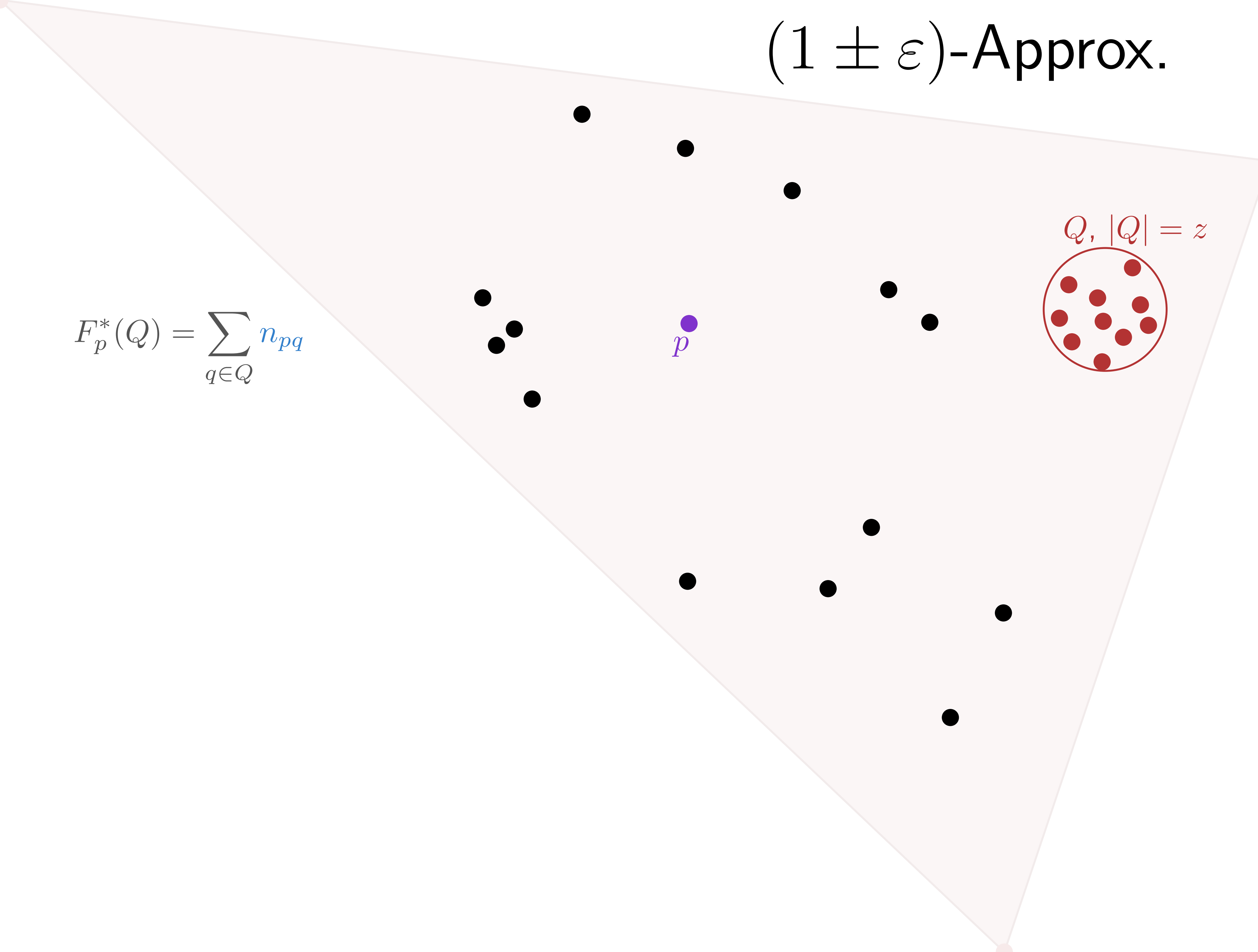
- Do not compute areas $\mathcal{A}'(\overrightarrow{pq})$ exactly

  Group pairs p, q that have approximately the same $\mathcal{A}'(\overrightarrow{pq})$

  $\Longrightarrow$ use a WSPD: $O(\frac{\rho^4}{\varepsilon^2} n \log n)$ pairs (p, Q) s.t. for all p and $q \in Q$ we have $(1-\varepsilon)\mathcal{A}'(\overrightarrow{pq}) \leq A_{pQ} \leq \mathcal{A}'(\overrightarrow{pq})$

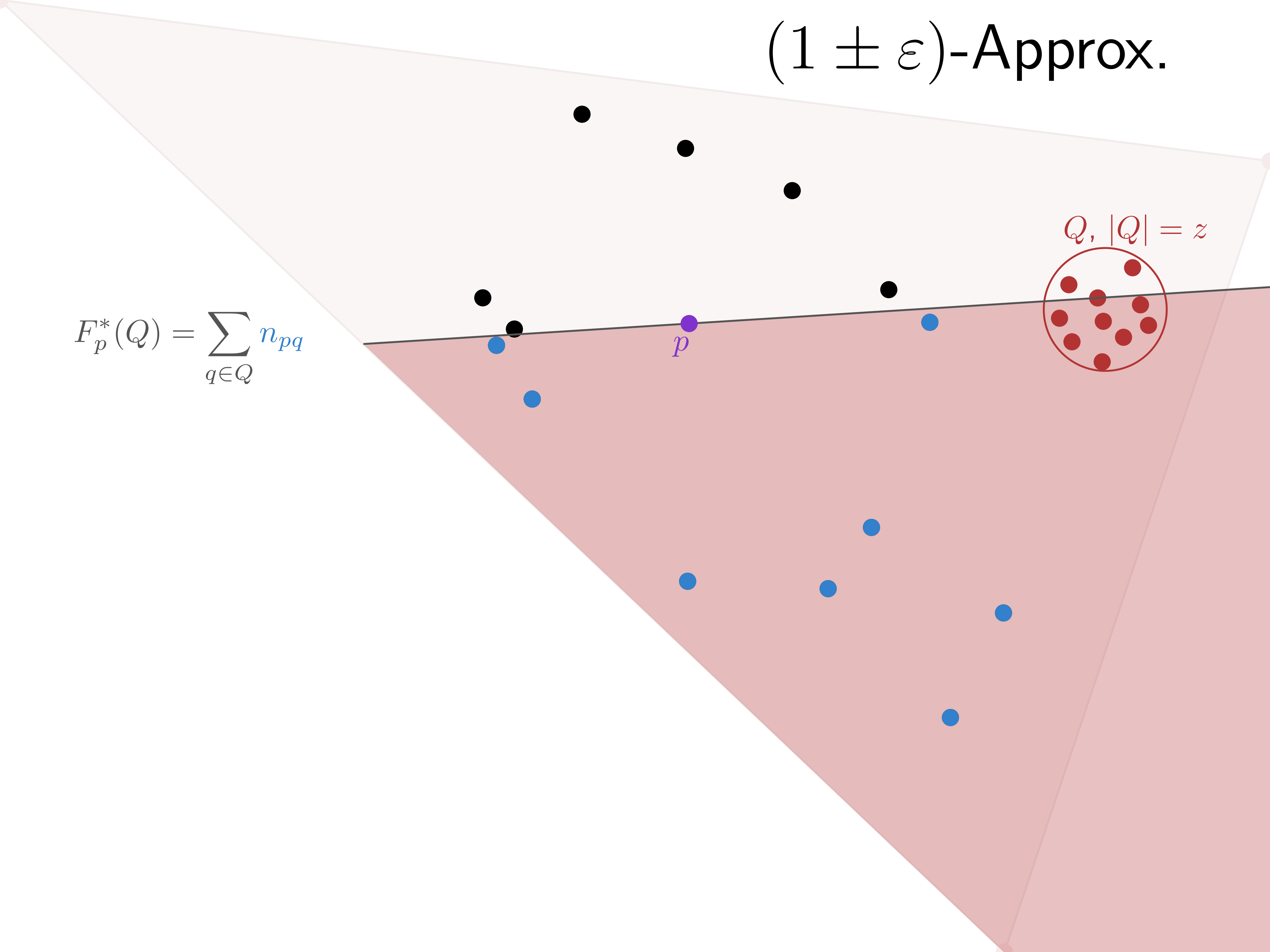
- Do not compute the counts n_{pq} exactly

$(1 \pm \varepsilon)$-Approx.

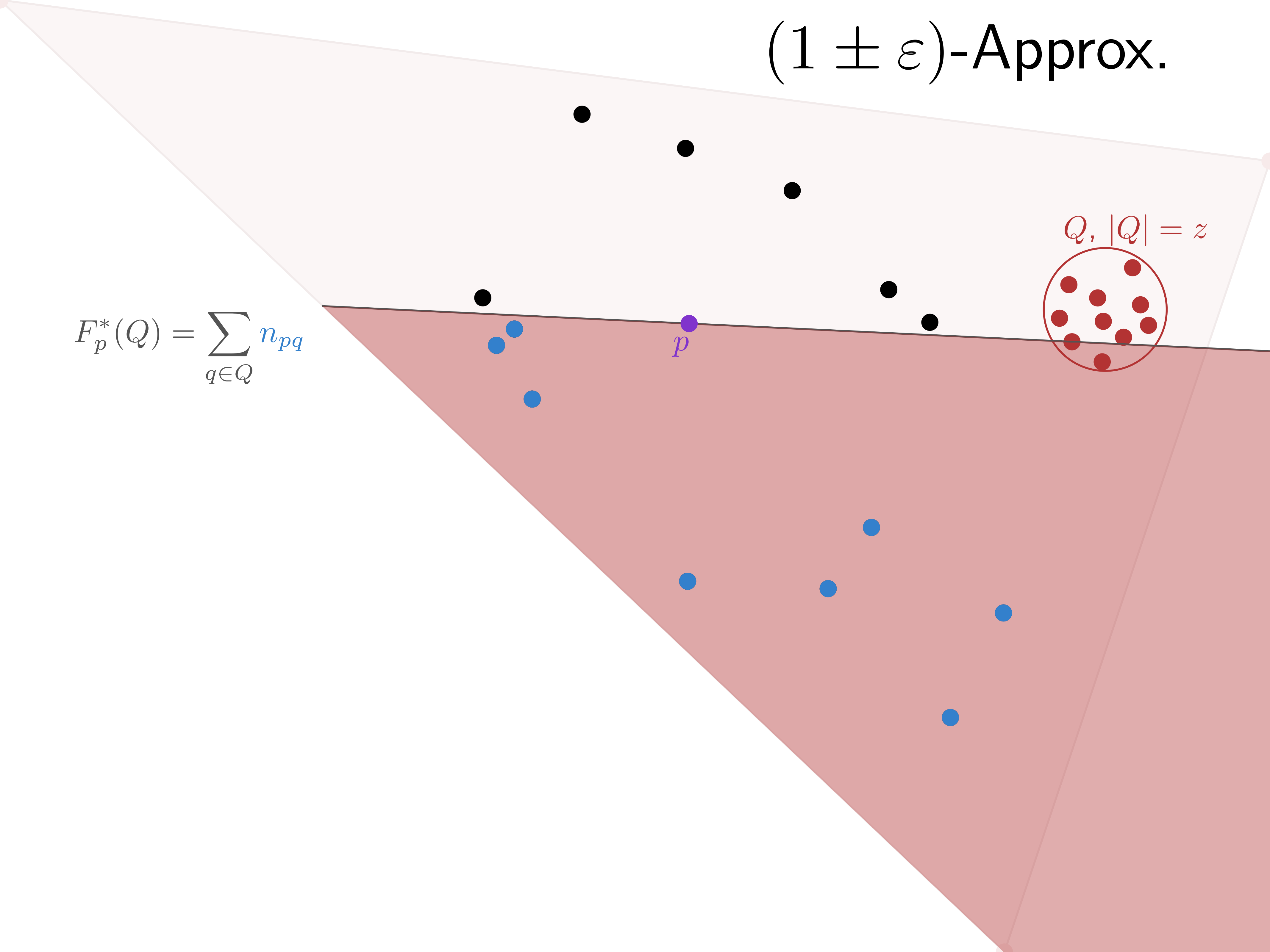


$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

$(1 \pm \varepsilon)$-Approx.



$(1 \pm \varepsilon)$-Approx.

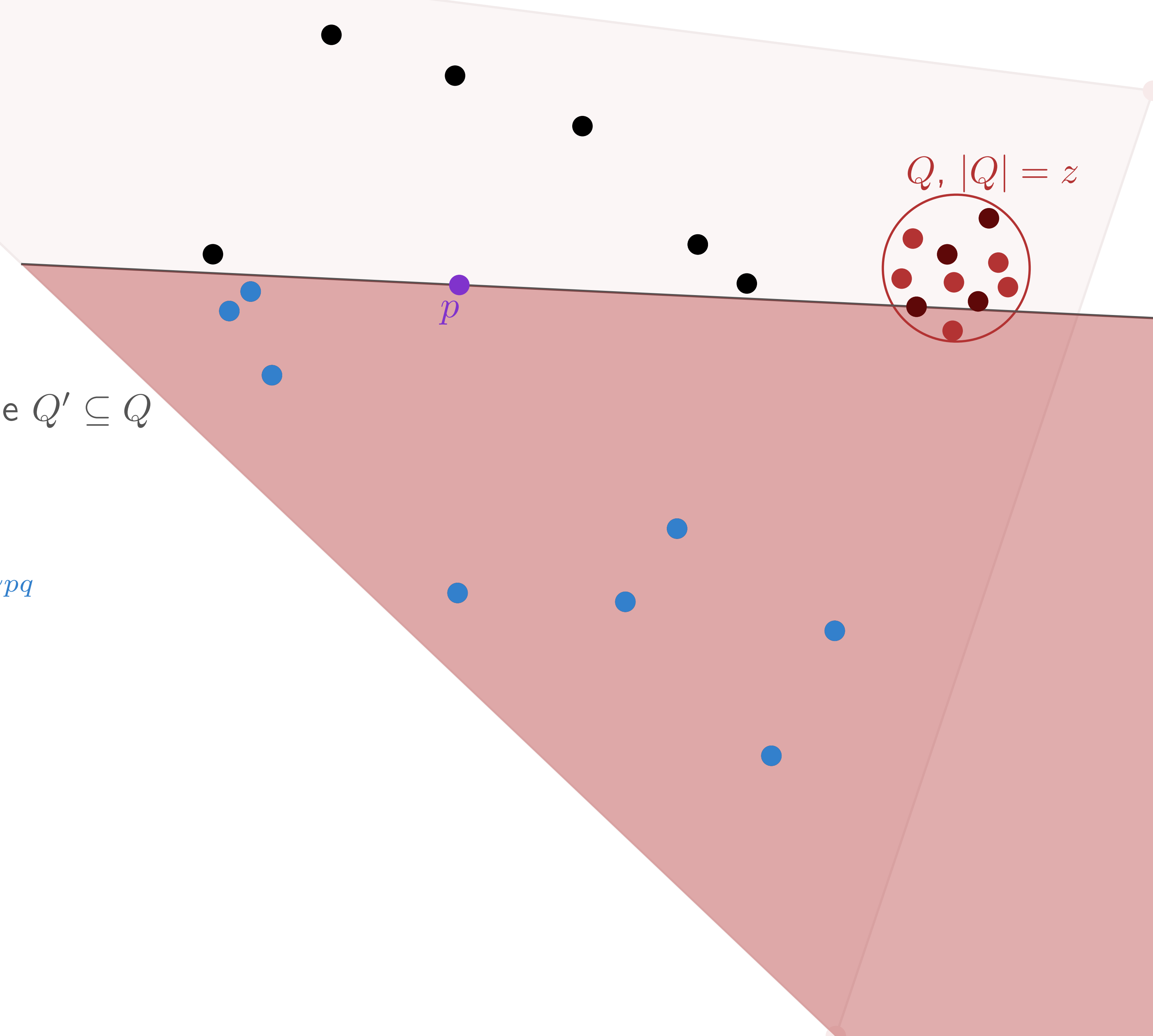


$(1 \pm \varepsilon)$-Approx.

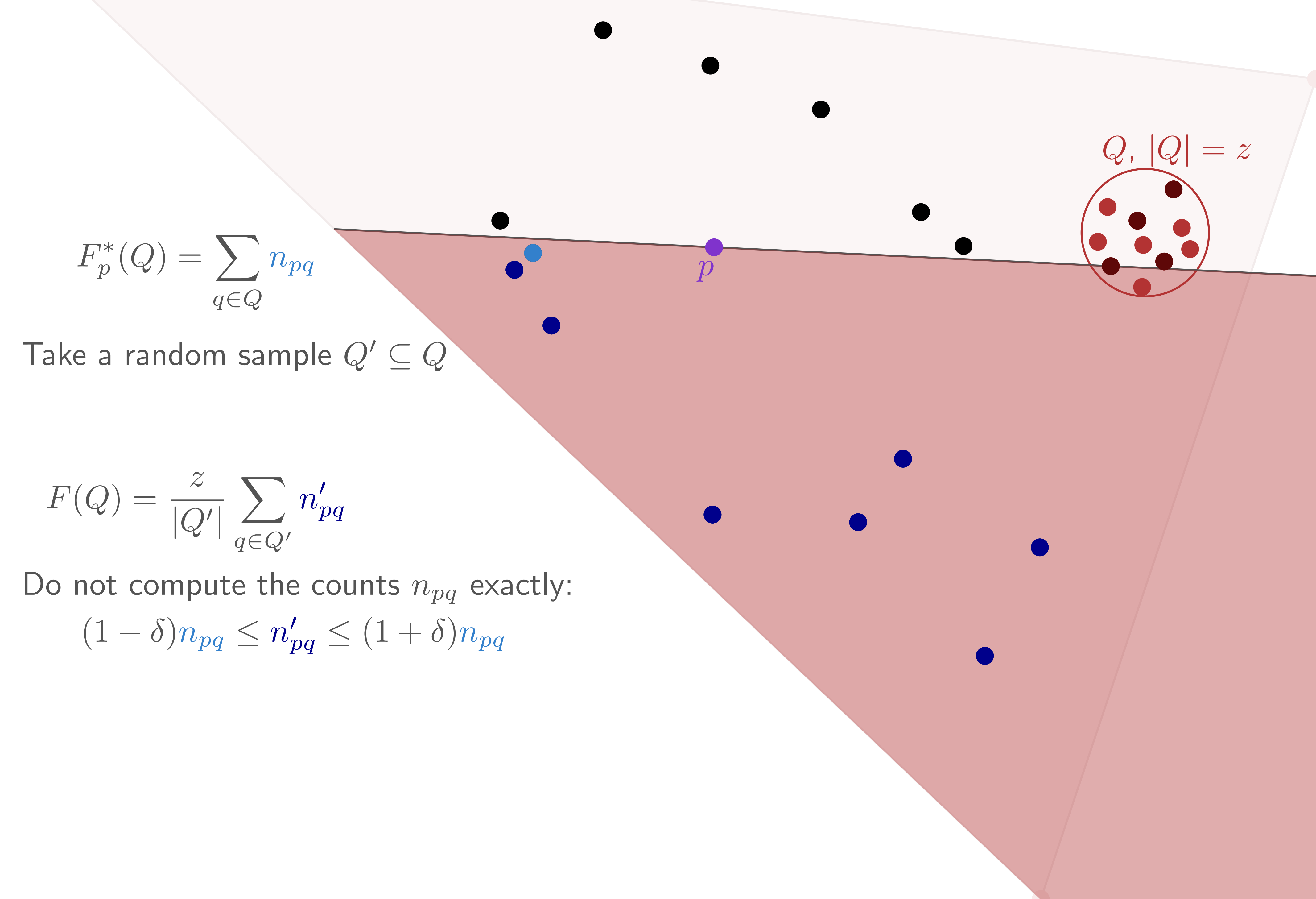
$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n_{pq}$$



$(1 \pm \varepsilon)$-Approx.



$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

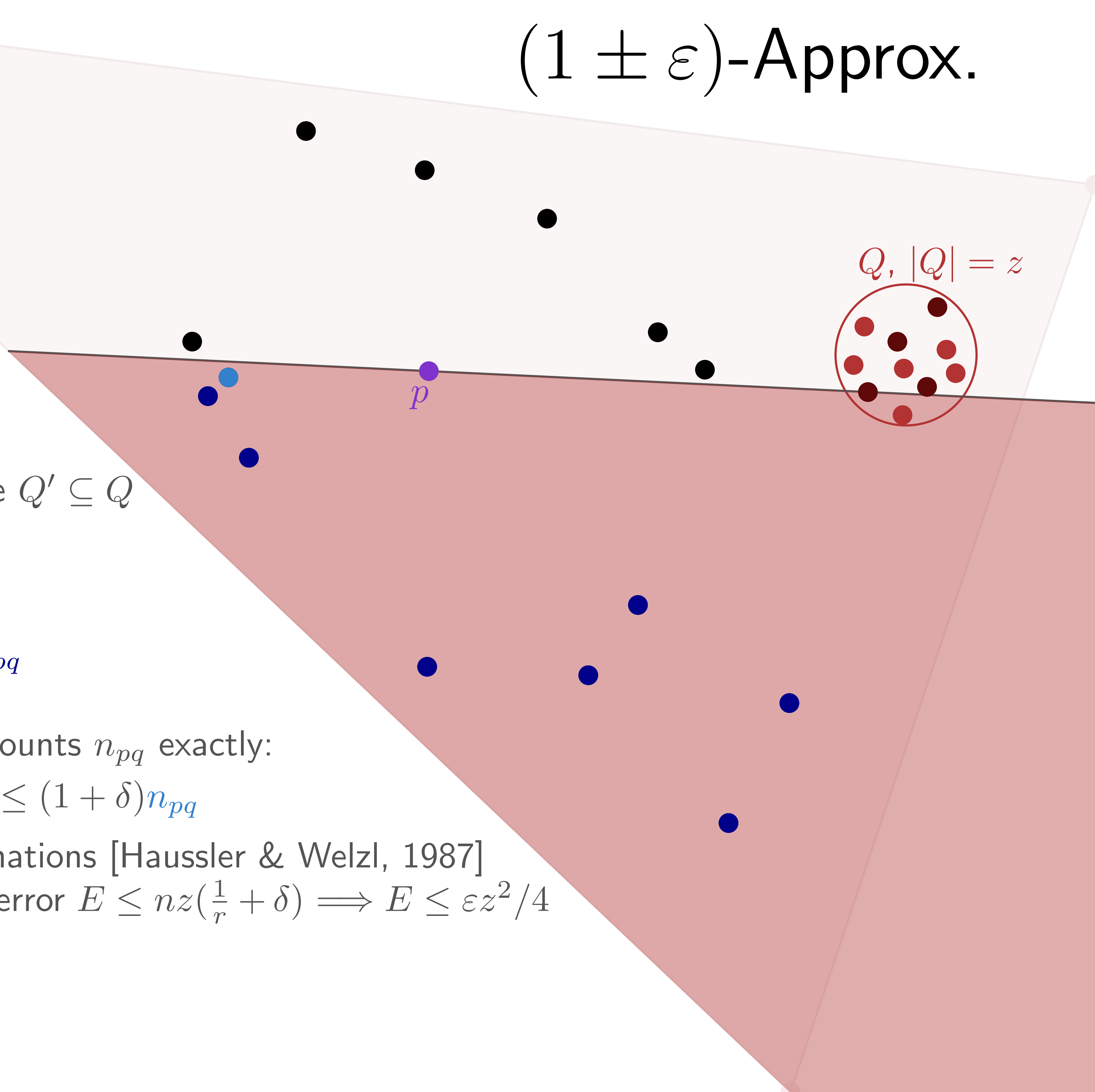
Take a random sample $Q' \subseteq Q$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

Do not compute the counts n_{pq} exactly:

$$(1-\delta) n_{pq} \leq n'_{pq} \leq (1+\delta) n_{pq}$$

$(1 \pm \varepsilon)$-Approx.



$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

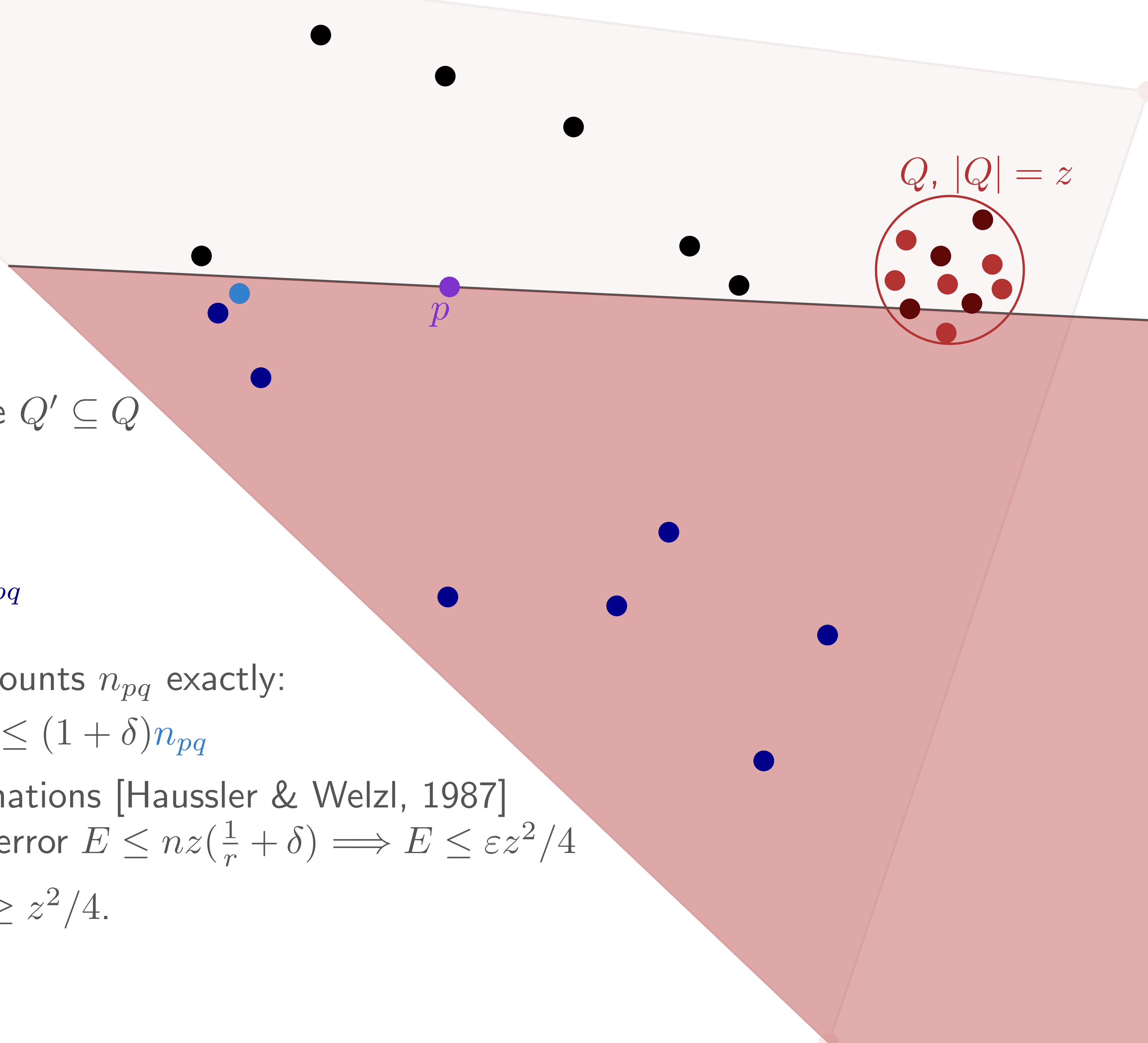
Do not compute the counts n_{pq} exactly:

$$(1 - \delta) n_{pq} \leq n'_{pq} \leq (1 + \delta) n_{pq}$$

$\Longrightarrow$ ε-nets/ε-approximations [Haussler & Welzl, 1987]

$\Longrightarrow$ .... $\Longrightarrow$ absolute error $E \leq nz(\frac{1}{r} + \delta) \Longrightarrow E \leq \varepsilon z^2/4$

$(1 \pm \varepsilon)$-Approx.



$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

Do not compute the counts n_{pq} exactly:

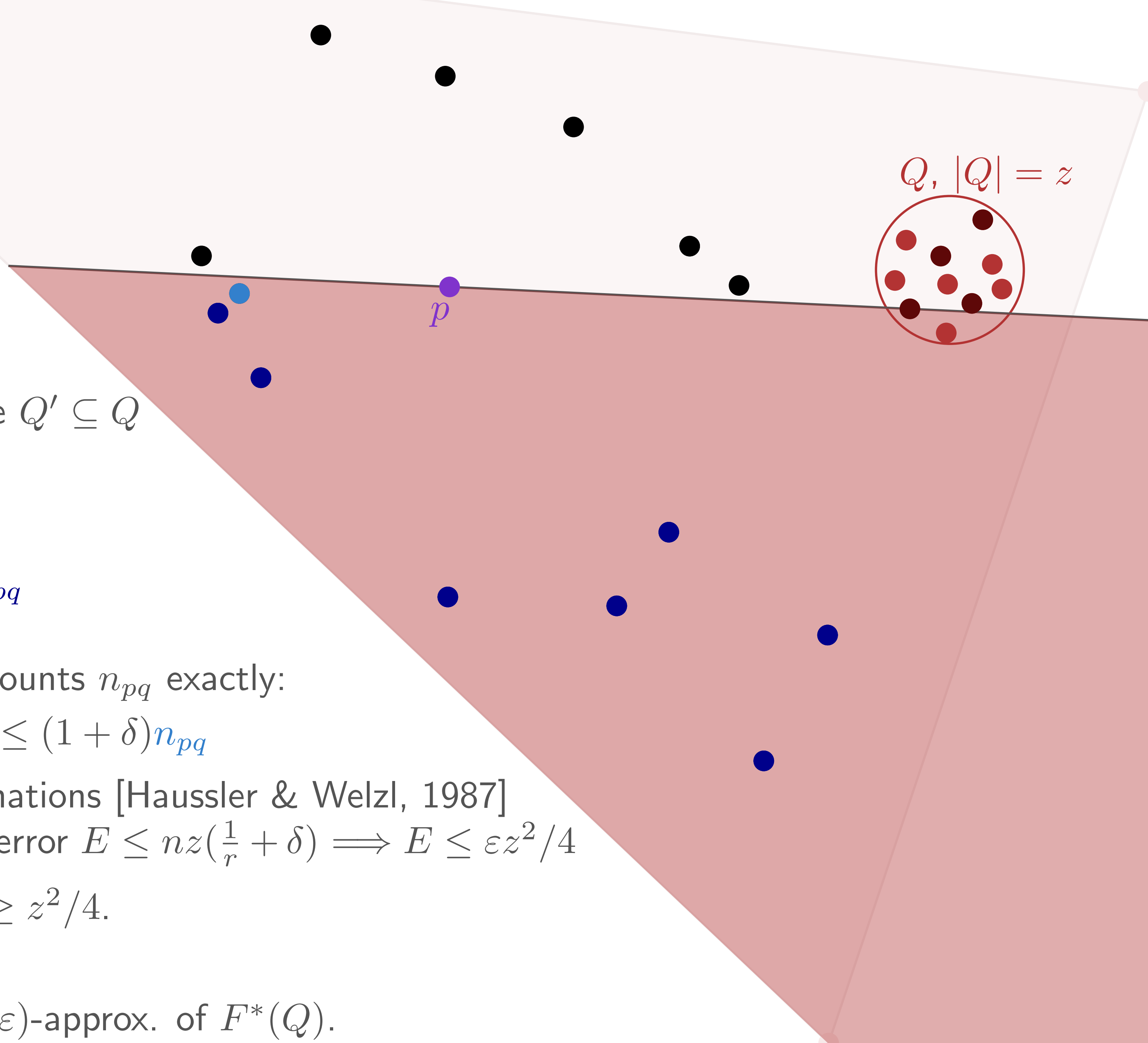
$$(1-\delta) n_{pq} \leq n'_{pq} \leq (1+\delta) n_{pq}$$

$\Longrightarrow$ ε-nets/ε-approximations [Haussler & Welzl, 1987]

$\Longrightarrow$ .... $\Longrightarrow$ absolute error $E \leq nz(\frac{1}{r} + \delta) \Longrightarrow E \leq \varepsilon z^2/4$

$F_p^*(Q) \geq z(z-1)/2 \geq z^2/4.$

$(1 \pm \varepsilon)$-Approx.



$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

Do not compute the counts n_{pq} exactly:

$$(1 - \delta) n_{pq} \leq n'_{pq} \leq (1 + \delta) n_{pq}$$

$\Longrightarrow$ ε-nets/ε-approximations [Haussler & Welzl, 1987]

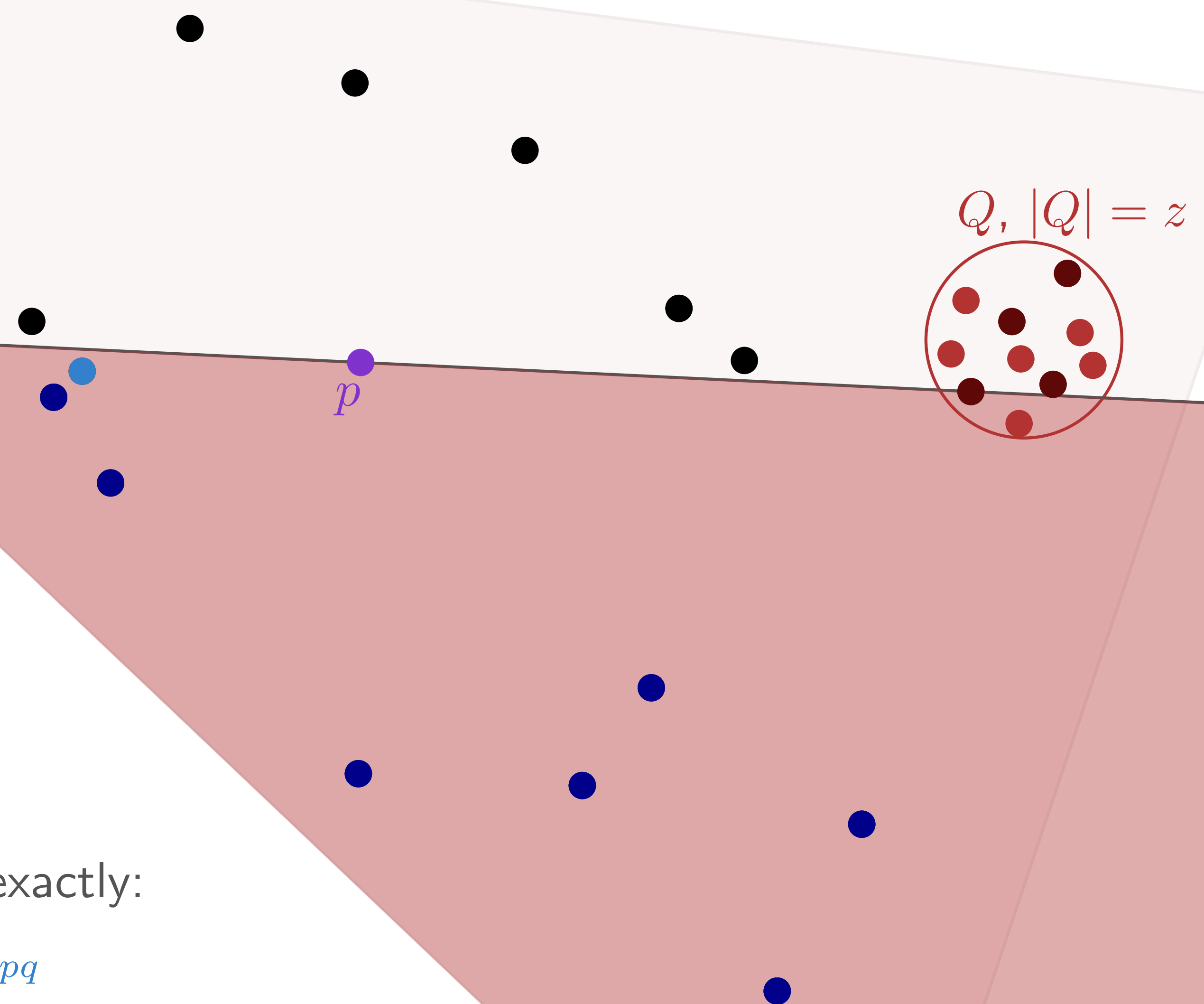
$\Longrightarrow$ .... $\Longrightarrow$ absolute error $E \leq nz(\frac{1}{r} + \delta) \Longrightarrow E \leq \varepsilon z^2/4$

$F_p^*(Q) \geq z(z-1)/2 \geq z^2/4$.

Lemma.

Whp. $F(Q)$ is a $(1 \pm \varepsilon)$-approx. of $F^*(Q)$.

$(1 \pm \varepsilon)$-Approx.



$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$
of size $\approx O((\frac{n}{\varepsilon})^{2/3})$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

Do not compute the counts n_{pq} exactly:

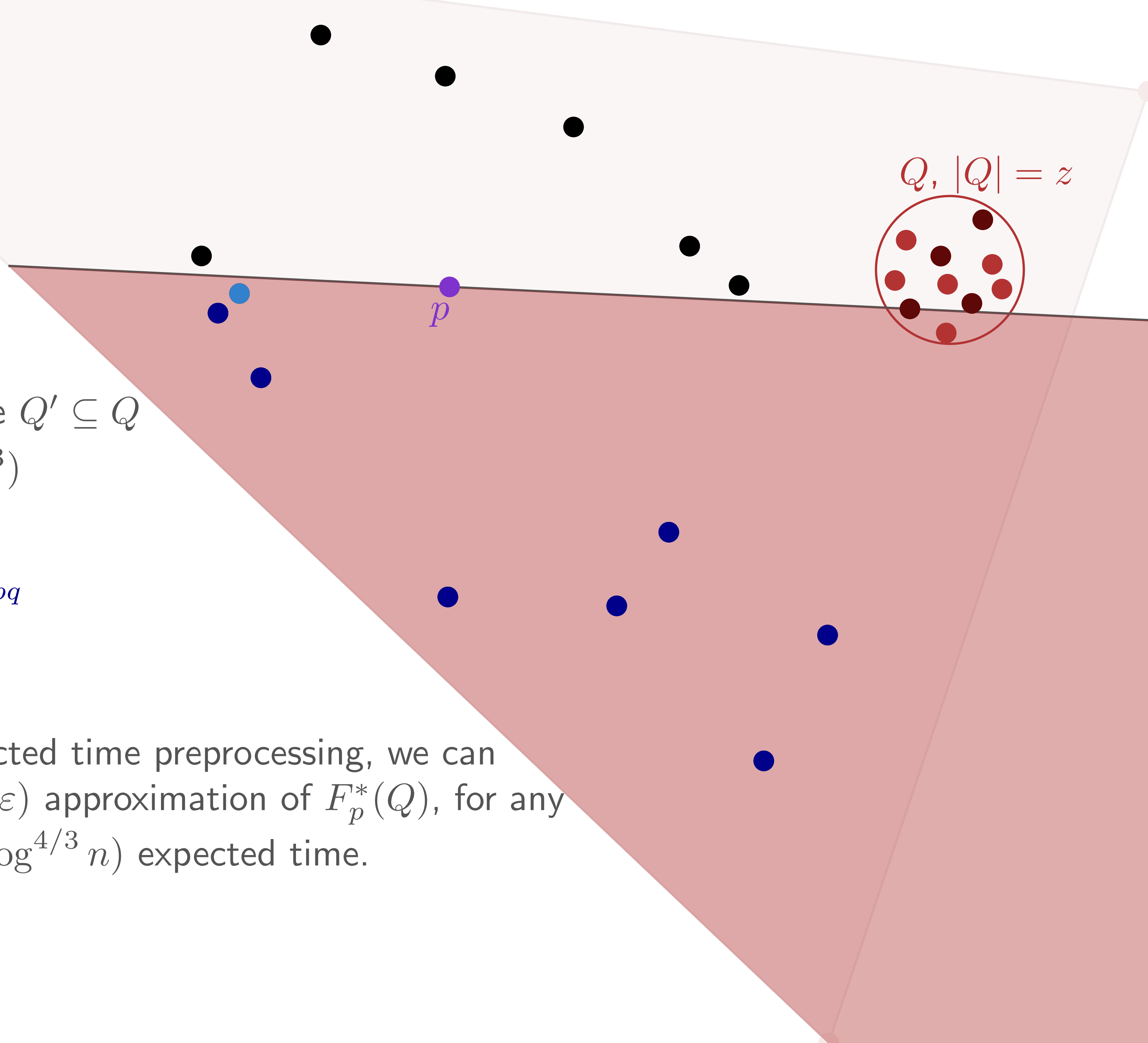
$$(1 - \delta) n_{pq} \leq n'_{pq} \leq (1 + \delta) n_{pq}$$

$\Longrightarrow$ ε-nets/ε-approximations [Haussler & Welzl, 1987]
$\Longrightarrow$ .... $\Longrightarrow$ absolute error $E \leq nz(\frac{1}{r} + \delta) \Longrightarrow E \leq \varepsilon z^2/4$

$F_p^*(Q) \geq z(z-1)/2 \geq z^2/4.$

Lemma.
Whp. $F(Q)$ is a $(1 \pm \varepsilon)$-approx. of $F^*(Q)$.

$(1 \pm \varepsilon)$-Approx.



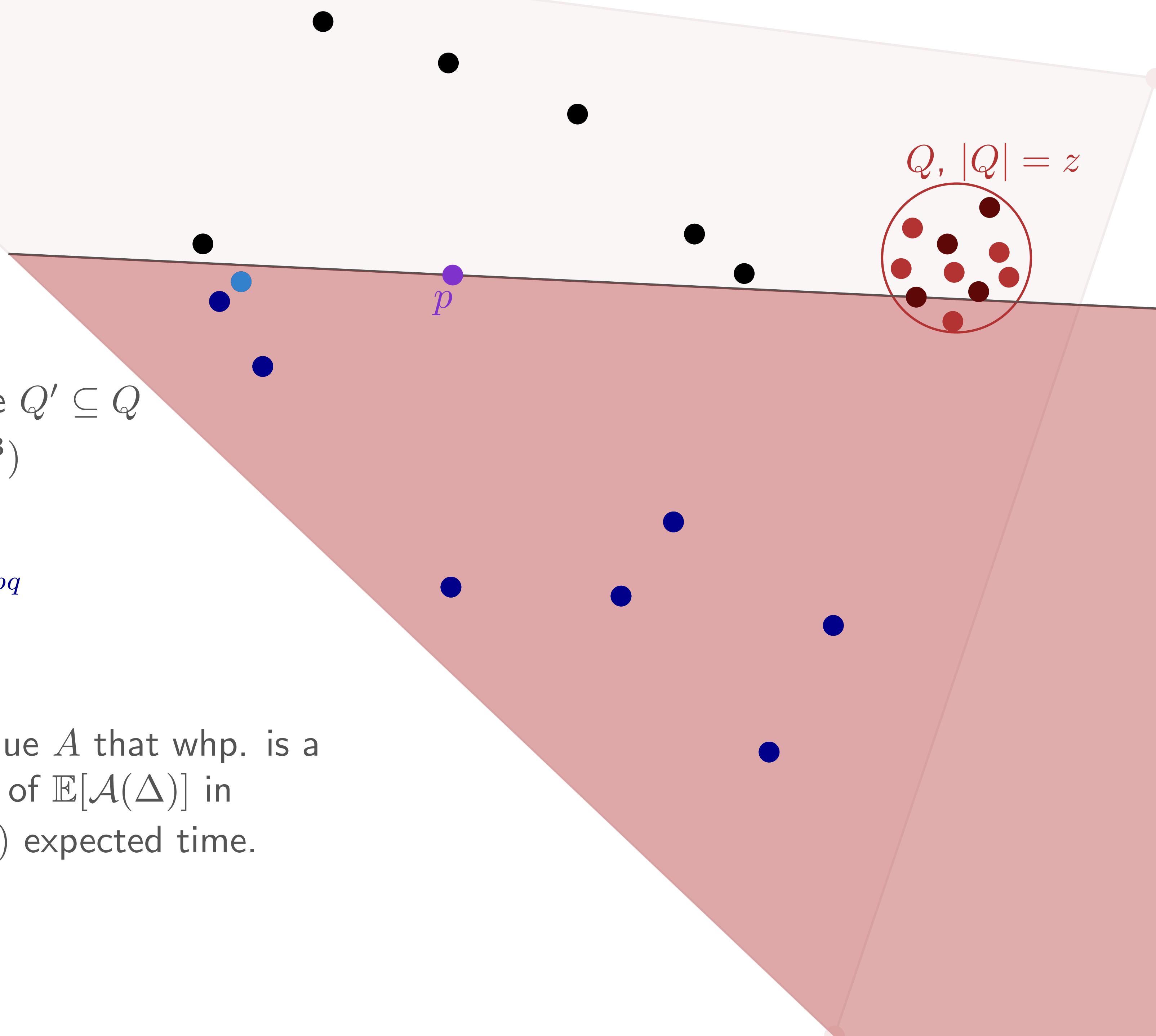
$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$
of size $\approx O((\frac{n}{\varepsilon})^{2/3})$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

Lemma.
After $O(n \log n)$ expected time preprocessing, we can whp. compute a $(1 \pm \varepsilon)$ approximation of $F_p^*(Q)$, for any (p, Q) in $O((n/\varepsilon)^{2/3} \log^{4/3} n)$ expected time.

$(1 \pm \varepsilon)$-Approx.



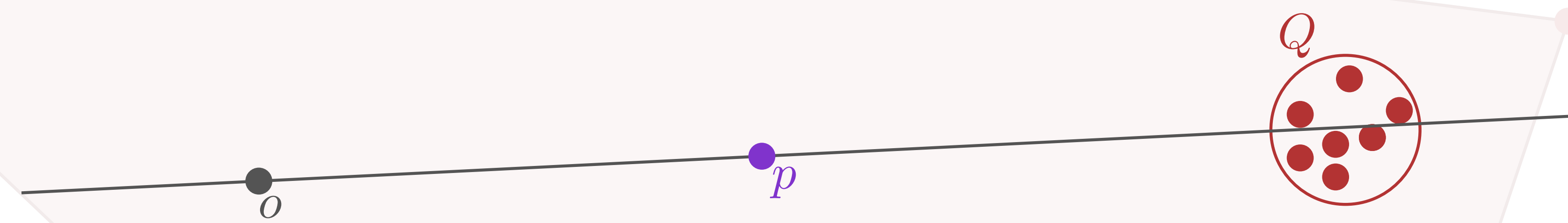
$$F_p^*(Q) = \sum_{q \in Q} n_{pq}$$

Take a random sample $Q' \subseteq Q$ of size $\approx O((\frac{n}{\varepsilon})^{2/3})$

$$F(Q) = \frac{z}{|Q'|} \sum_{q \in Q'} n'_{pq}$$

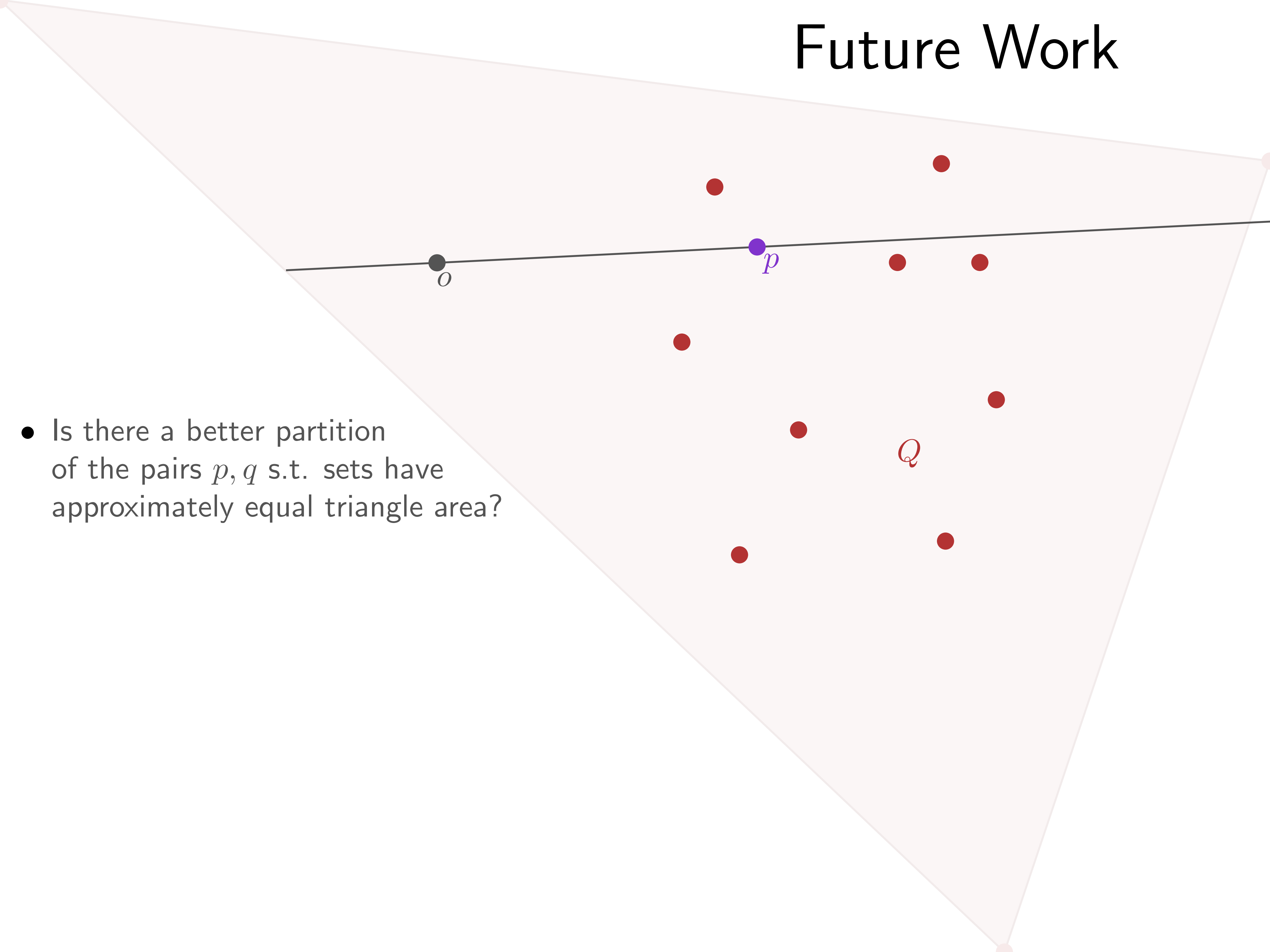
Theorem.
We can compute a value A that whp. is a $(1 \pm \varepsilon)$-approximation of $\mathbb{E}[\mathcal{A}(\Delta)]$ in $O(\frac{1}{\varepsilon^{8/3}} \rho^4 n^{5/3} \log^{7/3} n)$ expected time.

Future Work



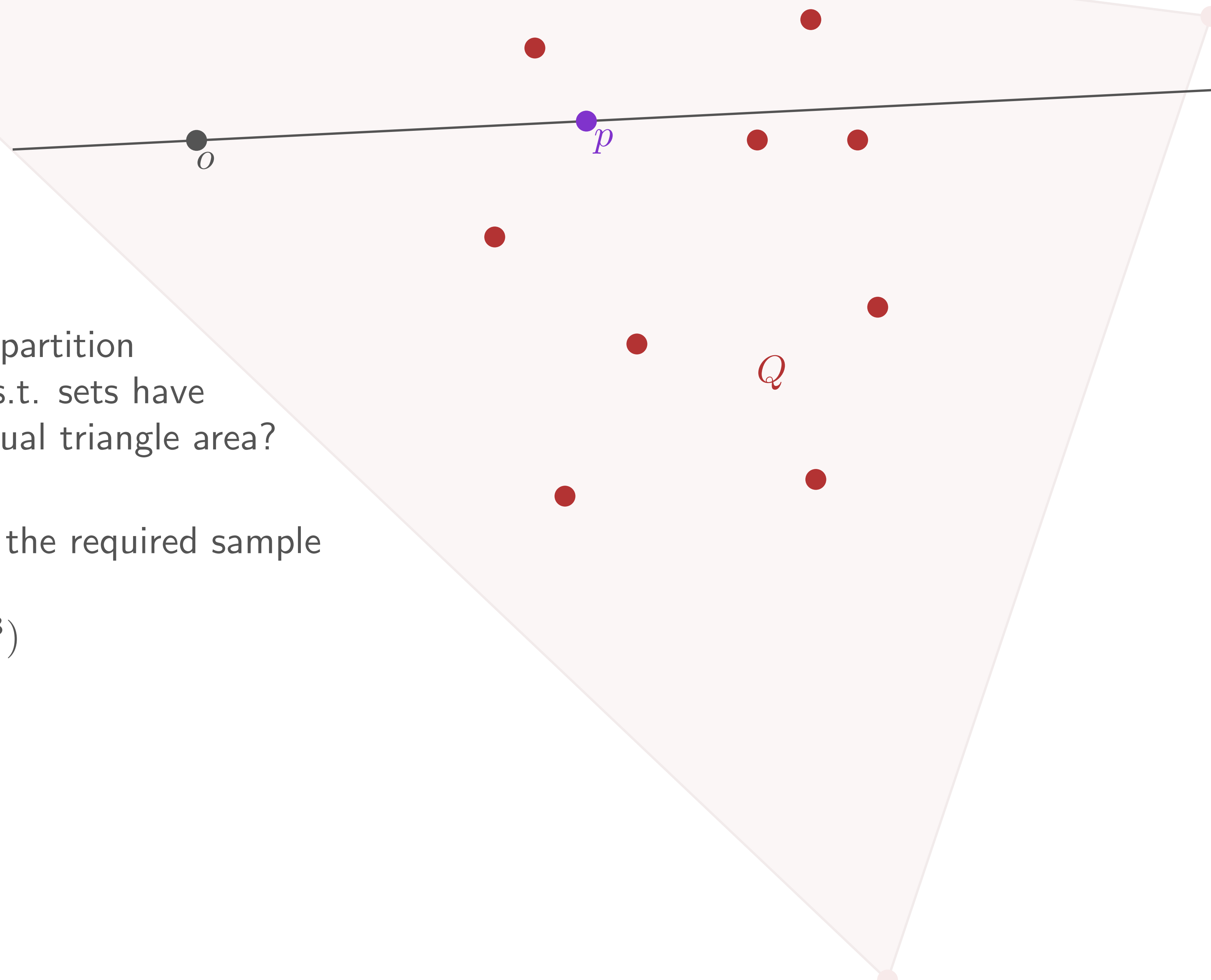
- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?

Future Work



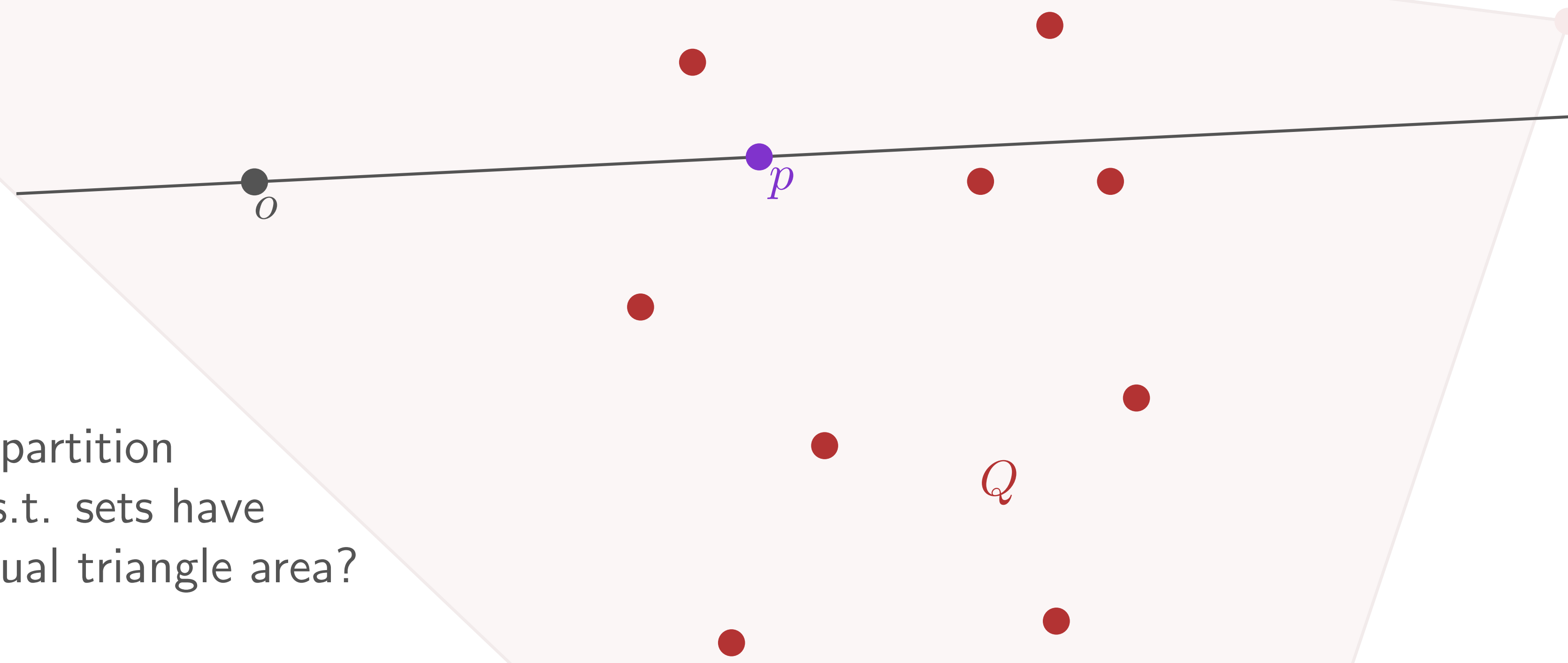
- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?

Future Work



- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?
- Can we decrease the required sample size?
  $|Q'| \approx O((\frac{n}{\varepsilon})^{2/3})$

Future Work

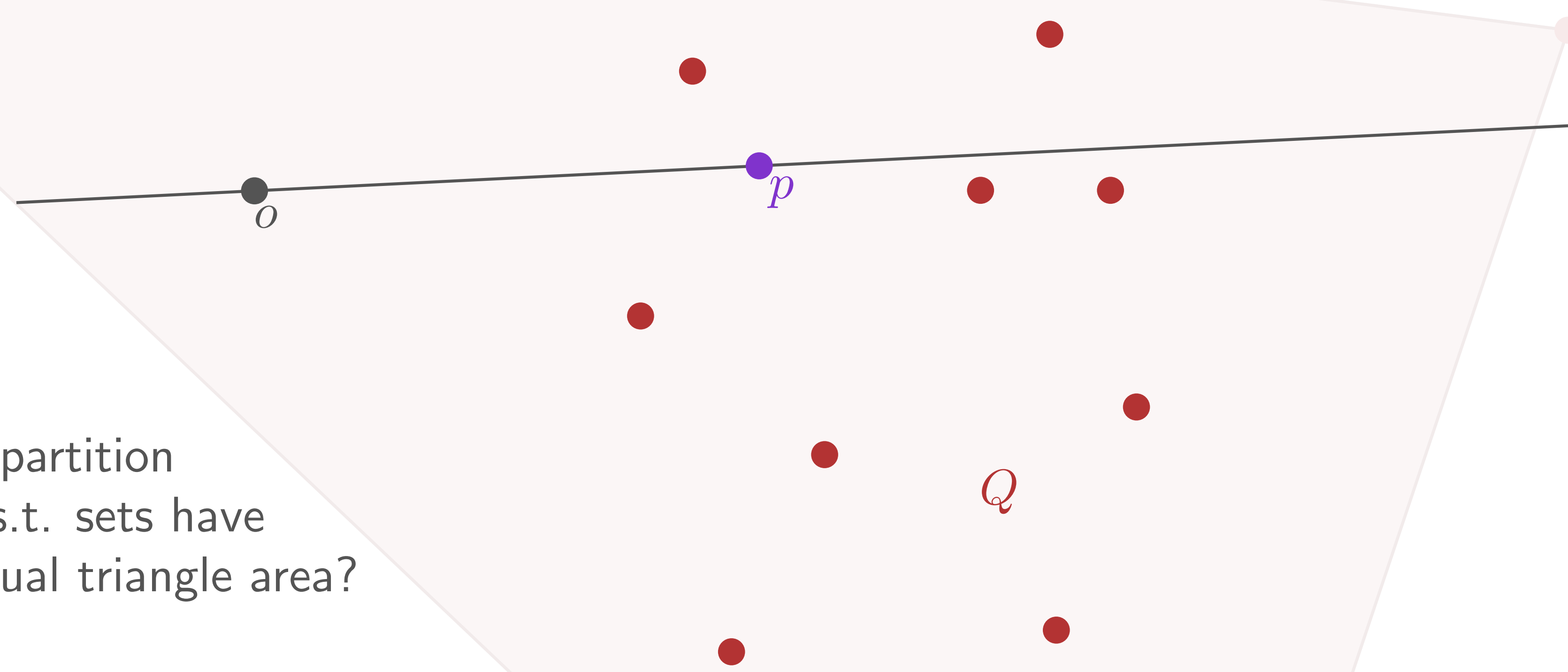


- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?

- Can we decrease the required sample size?
  $|Q'| \approx O((\frac{n}{\varepsilon})^{2/3})$

- Can we prove an $\Omega(n^2)$ lower bound for an exact solution?

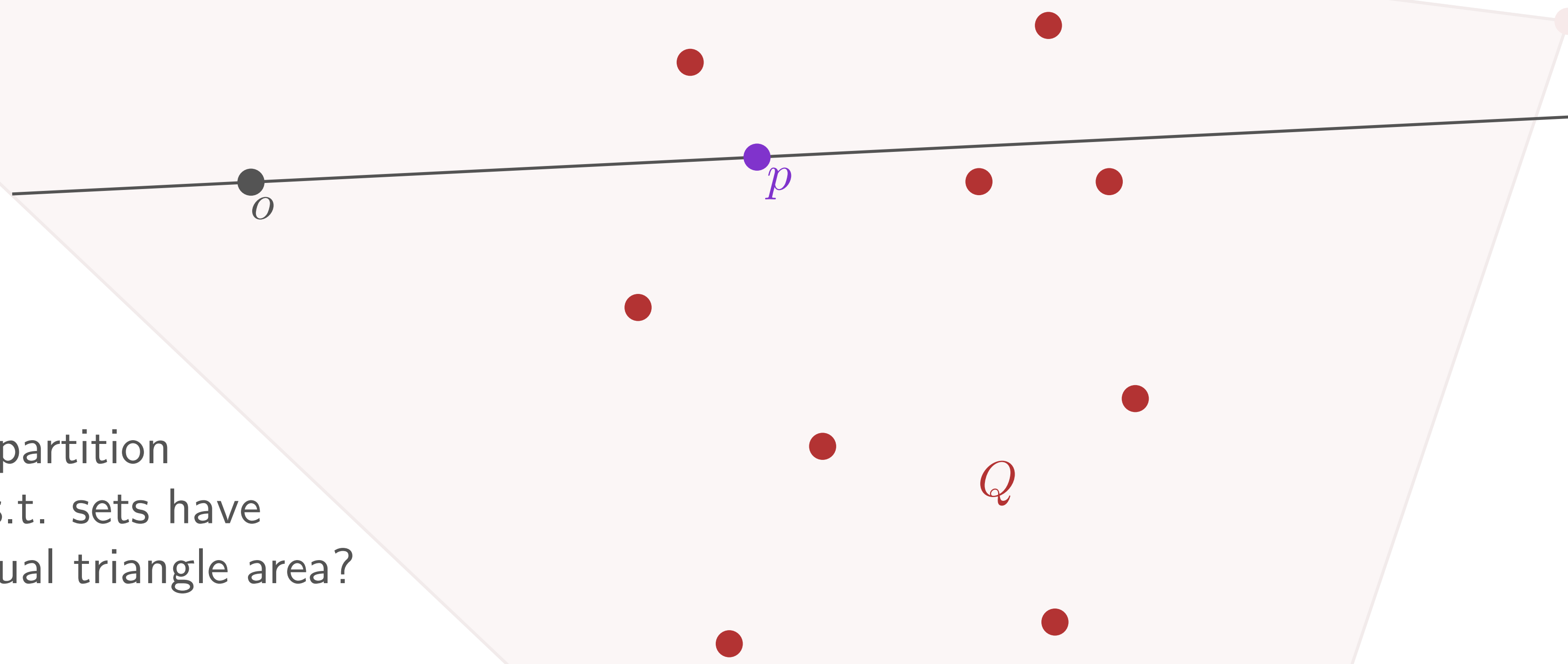
Future Work



- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?
- Can we decrease the required sample size?
  $|Q'| \approx O((\frac{n}{\varepsilon})^{2/3})$
- Can we prove an $\Omega(n^2)$ lower bound for an exact solution?
- $(1 \pm \varepsilon)$-approximation for convex hulls?

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \Delta] = \frac{\binom{n_q}{3-2}}{\binom{n}{3}} = \frac{n_q}{\binom{n}{3}} \qquad \Longrightarrow \sum n_{pq}$$

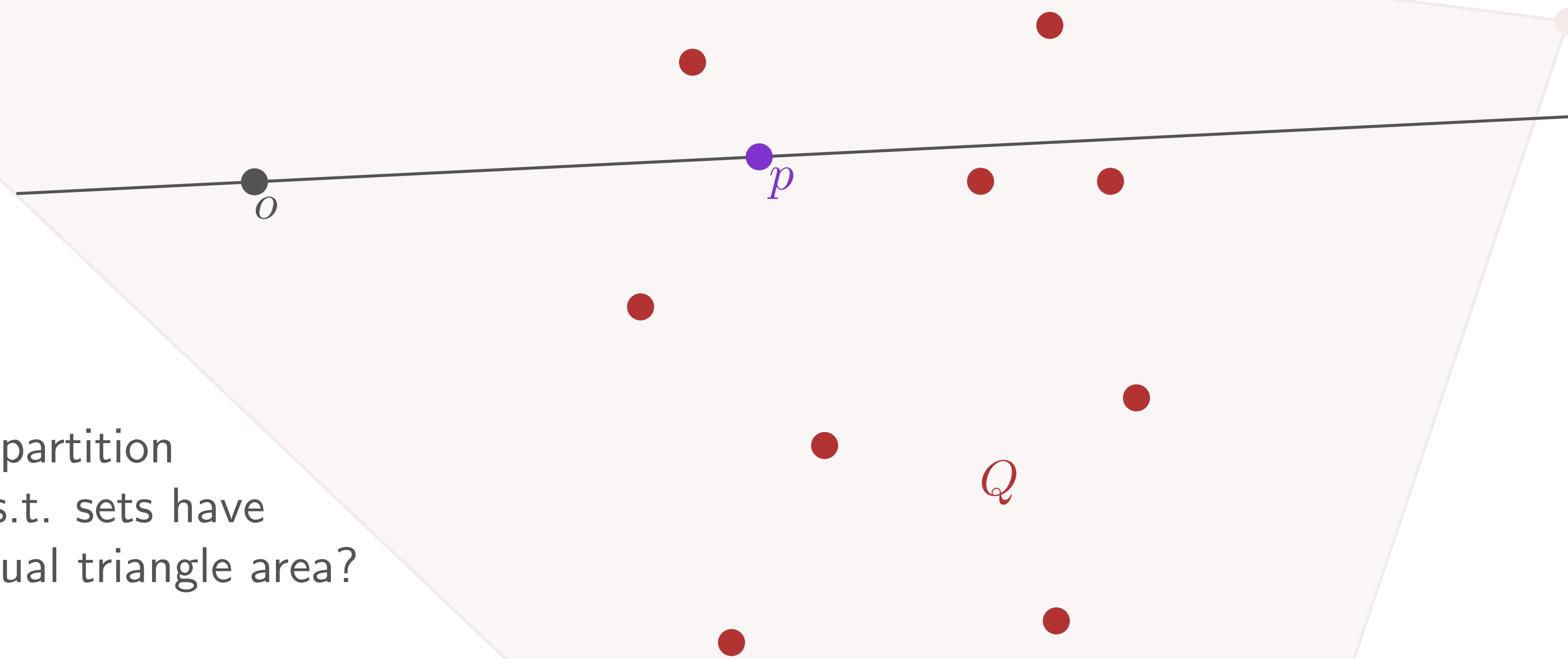
Future Work



- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?
- Can we decrease the required sample size?
  $|Q'| \approx O((\frac{n}{\varepsilon})^{2/3})$
- Can we prove an $\Omega(n^2)$ lower bound for an exact solution?
- $(1 \pm \varepsilon)$-approximation for convex hulls?

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \mathcal{CH}] = \frac{\binom{n_q}{k-2}}{\binom{n}{3}} = \frac{P_{k-2}(n_q)}{\binom{n}{3}} \implies \sum P_{k-2}(n_{pq})$$

Future Work



- Is there a better partition of the pairs p, q s.t. sets have approximately equal triangle area?

- Can we decrease the required sample size?
  $|Q'| \approx O((\frac{n}{\varepsilon})^{2/3})$

- Can we prove an $\Omega(n^2)$ lower bound for an exact solution?

- $(1 \pm \varepsilon)$-approximation for convex hulls?

$$\mathbb{P}[\overrightarrow{pq} \text{ is an edge of } \mathcal{CH}] = \frac{\binom{n_q}{k-2}}{\binom{n}{3}} = \frac{P_{k-2}(n_q)}{\binom{n}{3}} \implies \sum P_{k-2}(n_{pq})$$

Thank you!