

B) $q_i \leq q_{i+1}$.

We must show that $\frac{2^i}{p^2} \sum e_i^2 \leq \frac{2^{i+1}}{p^2} \sum e_{i+1}^2$.

The edges at the $(i+1)$ st stage are obtained by replacing each of the edges at the i th stage with two edges. We look at one of the edges in the i th stage, e_{ij} . Let it be replaced by $e_{i+1,2j-1}$ and $e_{i+1,2j}$. If we can show that $e_{ij}^2 \leq 2(e_{i+1,2j-1}^2 + e_{i+1,2j}^2)$, then by adding the above inequality over all edges of A_i we can obtain the equation above. The lengths of the edges of triangle T_{ij} are e_{ij} , $e_{i+1,2j-1}$, and $e_{i+1,2j}$. Hence, by Lemma 3.2.7, $e_{ij}^2 \leq 2(e_{i+1,2j-1}^2 + e_{i+1,2j}^2)$, so we are done. ■

Definition 3.2.13: Let $r_i = q_{i+1} - q_i$.

By the above claim, $\sum r_i \leq 1$ and $r_i \geq 0$. Notice that

$$\sum_{j=1}^{2^i} k_{ij} = \frac{p^2 r_i}{2^i}.$$

We bound the number of choices for B_{i+1} in terms of r_i . We do this by Kolmogorov complexity, i.e., by finding a short way of specifying B_{i+1} . We will first use B_i in the specification, and later recursively expand B_i .

Claim 3.2.14: The $(i+1)$ st approximation B_{i+1} of R can be specified without knowing R by giving the i th approximation B_i and two lists of numbers; a list of 2^i numbers between 1 and 9, and a list of 2^i numbers whose sum is at most $2^i(\alpha r_i \log n + \beta)$, where α and β are constants.

Proof: The approximation B_i tells us within $\sqrt{2}q_{i+1}$ where half the points of B_{i+1} are located. For every vertex in B_i , the corresponding vertex in B_{i+1} is either the same point or an adjacent point of the grid G_{i+1} . This is because these two points must both be the closest approximation to some fixed point on the boundary of R (See Figure 3.2.16). In the figure, a point in the marked square will be approximated by the point in the center of the square on grid G_i , and will be approximated by this point or one of the eight grid points on the

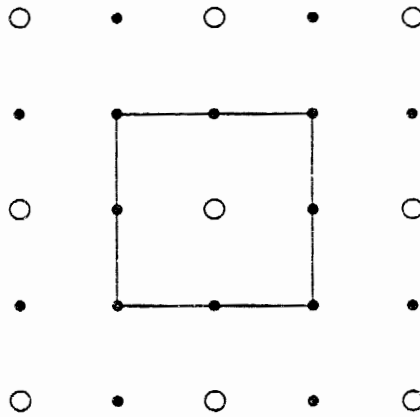


Figure 3.2.16: Approximations on grids G_{i+1} and G_i .

square on grid G_{i+1} . This gives 9 possible further approximation for each vertex of B_i . Thus, with a list of 2^i numbers between 1 and 9, we can specify for each of the 2^i points of B_i , which grid point of G_{i+1} it moves to. This list specifies half the vertices of B_{i+1} .

The remaining points of B_{i+1} are those which bisect an edge of B_i . For each edge of B_i , we label all the grid points of G_{i+1} , starting with the midpoint of the edge and using increasing labels with increasing distance from this midpoint. Since the edges of A_{i+1} have lengths at most $p/2^{i+1}$, the next vertex of B_{i+1} conceivably could be as far as $p/2^{i+1}$ away from this midpoint. This is $\sqrt{\log n}$ grid points of G_{i+1} away. However, most of the points of B_{i+1} will be considerably closer than this. To prove this, we need Lemma 3.2.6.

We use Lemma 3.2.6 to show that there are only a limited number of choices for B_{i+1} , given B_i and a bound on r_i . We do this by bounding the number of choices at each edge of B_i . Here we assume we are given a bound on $k_{ij} = 2(e_{i+1,2j-1}^2 + e_{i+1,2j}^2) - e_{ij}^2$, where e_{ij} is the length of the i th edge of A_i and $e_{i+1,2j-1}$ and $e_{i+1,2j}$ are the lengths of the edges of A_{i+1} which replace it. Since $r_i = \sum_{j=1}^{2^i} k_{ij}$, we will get the bound on r_i by summing all the bounds on k_{ij} .

This bound on the number of choices for a specific point of B_{i+1} follows from

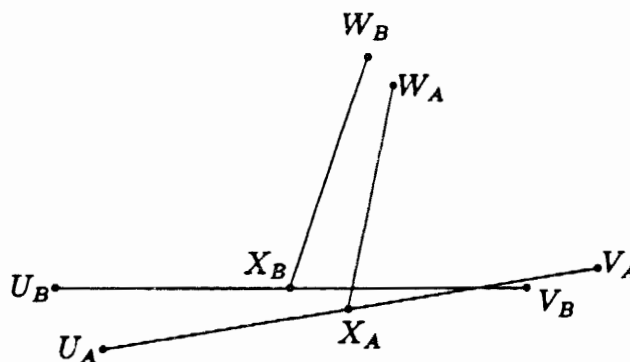


Figure 3.2.17:

Lemma 3.2.6 after some work. The idea is that given an edge e_{ij} of B_i , there is a circle around the midpoint of this edge of radius slightly more than $\frac{1}{2}\sqrt{k_{ij}}$ such that the vertex of B_i that splits edge e_{ij} will fall in this circle. Lemma 3.2.6 will help to obtain this circle.

Let U_A and V_A be the endpoints of the edge e_{ij} of A_i , and U_B and V_B be the endpoints of the corresponding edge of B_i . Let W_A and W_B be the vertices of A_{i+1} and B_{i+1} that divide the edge e_{ij} , and let X_A and X_B be the midpoints of the edge. (See Figure 3.2.17) Then, by Lemma 3.2.6, $d(X_A, W_A) \leq \frac{1}{2}\sqrt{k_{ij}}$. Since the vertices of B_i are approximations of the vertices of A_i on a grid with edge length g_i , $d(U_A, U_B) \leq g_i/\sqrt{2}$ and $d(V_A, V_B) \leq g_i/\sqrt{2}$. Since X_A is the midpoint of $U_A V_A$ and X_B of $U_B V_B$, this gives that $d(X_A, X_B) \leq g_i/\sqrt{2}$. We also have that $d(W_A, W_B) \leq g_i/2\sqrt{2}$, since W_B is an approximation to W_A in grid G_{i+1} . Using the triangle inequality, we get that $d(W_B, X_B) \leq \frac{1}{2}\sqrt{k_{ij}} + \frac{3}{4}\sqrt{2}g_i$. This is the desired inequality.

For each edge of B_i , we give a numbering to all the points of G_{i+1} , beginning with point 1 at the midpoint of the edge, and increasing the labels with increasing distance from the midpoint. With points equidistant from the midpoint we need to designate some arbitrary order, say starting with the topmost point and

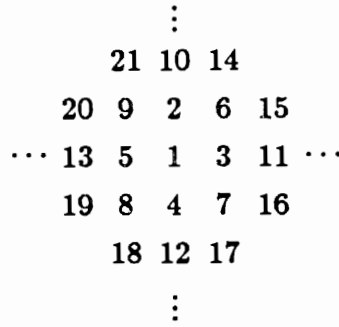


Figure 3.2.18: Numbering the grid points.

proceeding clockwise (See Figure 3.2.18). We let l_{ij} be the label of the point of B_{i+1} added on the j th edge of B_i .

We have showed that the point W of B_{i+1} will be in the circle of radius $r = \frac{1}{2}(\sqrt{k} + 3\sqrt{2}g_{i+1})$ centered on the midpoint of the edge. Since we labeled the points in increasing distance from the midpoint, the label l of the point W will be less than the number of grid points in that circle. However, by Lemma 3.2.4, the number of grid points in a circle of radius r is at most $\pi(r + g_{i+1}/\sqrt{2})^2/g_{i+1}^2$. Thus, the label of the point W is

$$\begin{aligned}
l &\leq \pi(r + \frac{1}{\sqrt{2}} g_{i+1})^2/g_{i+1}^2 \\
&= \pi(\sqrt{k}/2 + 2\sqrt{2}g_{i+1})^2/g_{i+1}^2 \\
&\leq \pi 2 \left((\sqrt{k}/2)^2 + (2\sqrt{2}g_{i+1})^2 \right) / g_{i+1}^2 \\
&= \frac{1}{2}\pi k/g_{i+1}^2 + 16\pi
\end{aligned}$$

We now show that the sum of the labels of points chosen for B_{i+1} satisfies

$$\sum_{j=1}^{2^i} l_{ij} \leq 2^i (\alpha r_i \log n + \beta)$$

for the constants $\alpha = \frac{\pi}{2}$ and $\beta = 16\pi$. Summing the inequality for l above, we get that

$$\sum_{j=1}^{2^i} l_{ij} \leq \left(\frac{1}{2}\pi/g_{i+1}^2 \right) \sum k_{ij} + 16\pi 2^i.$$

But recall our definition of k_{ij} was $2(e_{i,j-1}^2 + e_{ij}^2) - e_{i-1,2j}^2$. Thus,

$$r_i = \frac{2^{i+1}}{p^2} \sum e_{i+1}^2 - \frac{2^i}{p^2} \sum e_i^2,$$

so

$$r_i = \frac{2^i}{p^2} \sum k.$$

Also recall $g_{i+1} = g_1/2^i$. Thus,

$$\sum l \leq 2^i (p^2 \alpha / g_1^2) r_i + 2^i \beta.$$

However, from the definition of g_i , we have $g_1 \leq p/\sqrt{\log n}$. Thus, we get

$$\sum l \leq 2^i (\alpha r_i \log n + \beta).$$

This proves Claim 3.2.14. ■

We now use the above claim to bound the number of possible approximations B_{i+1} for any R . We do this by using Kolmogorov complexity. We show that we can determine B_1, B_2, \dots, B_{i+1} with two lists of numbers. We then bound the total number of possible lists of numbers and the total number of possible B_i 's. Since B_1, B_2, \dots, B_i determine R_1, R_2, \dots, R_i , this gives the bound on the number of possible R_i in the following claim.

Claim 3.2.15: The number of possible sequences R_1, R_2, \dots, R_{i+1} for all R with

$$r_i + \frac{r_{i-1}}{2} + \frac{r_{i-2}}{4} + \dots + \frac{r_1}{2^{i-1}} \leq r'$$

is bounded by

$$n 3_2 2^{i+1} \log \alpha'' r'_i \log n + \beta''$$

for some constants α'' and β'' .

Applying claim 3.2.14 recursively, we see that B_{i+1} can be specified by B_1 , 2^{i+1} numbers between 1 and 9 and 2^{i+1} numbers whose sum is

$$2^i \left(\alpha \left(r_i + \frac{1}{2} r_{i-1} + \frac{1}{4} r_{i-2} + \dots + \frac{1}{2^{i-1}} r_1 \right) \log n + 2\beta \right).$$

Let

$$r'_i = r_i + \frac{1}{2}r_{i-1} + \frac{1}{4}r_{i-2} + \dots + \frac{1}{2^{i-1}}r_1.$$

Then the sum of the numbers l is at most

$$2^i (\alpha r'_i \log n + 2\beta).$$

Since $\sum r_i \leq 1$, we get

$$\sum r'_i \leq \sum r_i (1 + \frac{1}{2} + \frac{1}{4} + \dots) \leq 2.$$

We can determine B_{i+1} with B_1 , and two lists of integers: a list of 2^{i+1} integers each of which is between 1 and 9, and a list of 2^{i+1} positive integers whose sum is bounded by

$$2^i (p^2 \alpha / g_1^2) r'_i + 2^{i+1} \beta.$$

The number of ways of making the first list of 2^{i+1} integers is $9^{2^{i+1}}$. The number of ways of making the second list is

$$\binom{2^i (\alpha r'_i \log n + 2\beta)}{2^{i+1}}.$$

Using the inequality

$$\binom{ab}{b} \leq (ea)^b,$$

we get the bounds

$$\begin{aligned} \binom{2^i (\alpha r'_i \log n + 2\beta)}{2^{i+1}} &\leq \left(e \left(\frac{1}{2} \alpha r'_i \log n + \beta \right) \right)^{2^{i+1}} \\ &= 2^{2^{i+1} \log(\alpha' r'_i \log n + \beta')}. \end{aligned}$$

The total number of choices for the sequence B_1, B_2, \dots, B_{i+1} , given B_1 is thus at most

$$\begin{aligned} 9^{2^{i+1}} 2^{2^{i+1} \log(\alpha' r'_i \log n + \beta')} &= 2^{2^{i+1} (\log 9 + \log(\alpha' r'_i \log n + \beta'))} \\ &= 2^{2^{i+1} \log(\alpha'' r'_i \log n + \beta'')}. \end{aligned}$$

We must still bound the number of choices for B_1 . This is easy, since there are only 2 points in B_1 (i.e., it is a line segment). We bound B_1 by looking at all possible grid lengths for G_1 and all possible pairs of points of G_1 . The grid length of G_1 depends only on the perimeter (and n). As we are taking the perimeter to be a power of 2, and as the maximum possible perimeter is n^2 , there are at most $\log(n^2) = 2\log n$ possibilities for the grid length. The grid length will always be at least 1, so there are at most n points on the grid. Thus, the number of ways of choosing 2 of them to obtain B_1 is at most n^2 . We thus have that there are at most $2n^2 \log n \leq n^3$ possibilities for B_1 . This shows that the number of possibilities for B_{i+1} given n and a bound on r'_i (and not given R) is at most

$$n^3 2^{i+1} \log(\alpha'' r'_i \log n + \beta''),$$

proving Claim 3.2.15.

We have bounded the number of choices for $R_{i+1} - R_i$; we must also bound the area of $|R_{i+1} - R_i|$. Intuitively, we look at the triangles T_{ij} along the border of A_i . The union of these triangles is approximately $|R_{i+1} - R_i|$. Since the average angle of these triangles is small, their average area will also be small. What we will actually do is to take the union of these triangles to get C_i . We then take everything within $\sqrt{2}g_{i+1}$ of C_i to get D_i , and bound the area of D_i . We then show that $R_{i+1} - R_i \subseteq D_i$.

Claim 3.2.16: The area of R_i is at most $p^2 2^{-i} \sqrt{2r_i + \gamma / \log n}$ for some constant γ .

Proof: We first need to define two regions which we will use to prove the bound. They will be called C_i and D_i . The region C_i will be the union of all the triangles forming the difference of A_{i+1} and A_i . The region D_i will be the region formed by taking all points within $\sqrt{2}g_{i+1}$ of C_i . We obtain a bound on the area of D_i and then show that $|R_{i+1} - R_i| \subseteq D_i$.

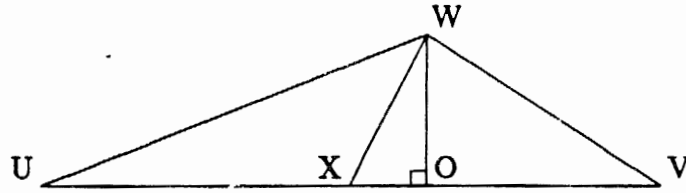


Figure 3.2.19: The triangle T_{ij} .

Definition 3.2.17: The region C_i is $\bigcup_{j=1}^{2^i} T_{ij}$, where T_{ij} is the triangle with vertices $a_{i,j-1}$, a_{ij} and $a_{i+1,2j-1}$.

Definition 3.2.18: The region D_i is the set of all points within $\sqrt{2}g_{i+1}$ of a point in C_i .

The area of C_i is bounded by the sum of the areas of all the triangles T_{ij} , since one obtains C_i by taking the union of these triangles. The altitude of triangle T_{ij} , however, is at most the length of the median, which by Lemma 3.2.6 is $\frac{1}{2}\sqrt{k_{ij}}$. In Figure 3.2.19, this corresponds to the inequality $|WO| \leq |WX|$. Thus, the triangle T_{ij} will have area at most $\sqrt{k_{ij}}e_{ij}/4$, where e_{ij} the length of the j th edge of A_i . We have that $e_{ij} \leq p/2^i$, since the distance along the perimeter between the endpoints of e_{ij} is $p/2^i$. Thus, $\text{Area}(T_{ij}) \leq \sqrt{k_{ij}}p/2^i$. Adding this inequality over all triangles, one obtains a bound of

$$\text{Area}(C_i) \leq \sum_j \text{Area}(T_{ij}) \leq \frac{p}{2^i} \sum_{j=1}^{2^i} \sqrt{k_{ij}}.$$

However,

$$\sum_{j=1}^{2^i} \sqrt{k_{ij}} \leq 2^i \sqrt{\frac{\sum_{j=1}^{2^i} k_{ij}}{2^i}} = p\sqrt{r_i}$$

since $\sum k_{ij} = p^2 r_i / 2^i$. The bound on the area of C_i then becomes

$$\text{Area}(C_i) \leq \frac{p^2 \sqrt{r_i}}{2^i}.$$

We now show that the area of D_i is at most

$$\frac{p^2 \sqrt{r_i}}{2^i} + \sqrt{2}pg_i + \pi g_i^2/2$$

We use Lemma 3.2.5. Recall this lemma gives a bound of $2ld + \pi d^2$ on the area of the region within distance d of a path of length l . The perimeter of C_i is at most $2p$, since

$$\sum_{j=1}^{2^i} \text{Per}(T_{ij}) = \text{Per}(A_i) + \text{Per}(A_{i+1}) \leq 2p.$$

The boundary of C_i is contained in the union of the polygons A_i and A_{i+1} , i.e.

$$\partial C_i \subseteq A_i \cup A_{i+1}.$$

The union of A_i and A_{i+1} clearly forms a path in the plane. Thus, by the Lemma 3.2.5, we have that the area of everything within $\sqrt{2}g_{i+1}$ of the boundary of C_i is at most $\pi(\sqrt{2}g_{i+1})^2 + 2p\sqrt{2}g_{i+1}$. Adding this to our bound for $\text{Area}(C_i)$, we obtain a bound for $\text{Area}(D_i)$ of

$$\frac{p^2 \sqrt{r_i}}{2^i} + \sqrt{2}pg_i + \pi g_i^2/2.$$

We wish to change this into a more tractable form. Since $p > g_i$, we have $\sqrt{2}pg_i + \pi g_i^2/2 \leq c_1 pg_i$ for some constant c_1 . Now, $g_i = \Theta(p2^{-i}/\sqrt{\log n})$. Using this, we have

$$\begin{aligned} \text{Area}(D_i) &\leq \frac{p^2}{2^i} \left(\sqrt{r_i} + \frac{c_2}{\sqrt{\log n}} \right) \\ &\leq \frac{p^2}{2^i} \sqrt{2r_i + 2c_2^2/\log n}. \end{aligned}$$

With $\gamma = 2c_2^2$, this is the desired bound for $\text{Area}(|R_{i+1} - R_i|)$.

We must now show that $|R_{i+1} - R_i| \subseteq D_i$. We will do this by showing that every point in the square either has the same winding number with respect to the curves A_i and B_i or is within $g_i/\sqrt{2}$ of the curve A_i . This proves that $|R_{i+1} - R_i| \subseteq D_i$ because a point x is in $|R_{i+1} - R_i|$ only if it has a different winding number with respect to B_i and B_{i+1} . This shows that x either has a different winding number with respect to A_i and A_{i+1} or that it is within $g_i/\sqrt{2}$ of one of A_i and A_{i+1} . If x has different winding numbers for A_i and A_{i+1} , then it is in one of the triangles T_{ij} , and so is in C_i . If x is within $g_i/\sqrt{2}$ of a point on A_i or A_{i+1} , then it is within $g_i/\sqrt{2}$ of C_i and so is in D_i .

The polygon B_i is produced by moving all the vertices of the polygon A_i so they lie on grid points. We never move any of these vertices more than $g_i/\sqrt{2}$. Thus, we never move any of the edges more than $g_i/\sqrt{2}$. A point x will have the same winding number with respect to A_i and B_i unless it is on one side of an edge of A_i and on the other side of the corresponding edge of B_i . This will not happen unless an edge of A_i was moved across x , which cannot happen if the point x is farther than $g_i/\sqrt{2}$ from A_i . ■

We have now shown that the area of $|R_{i+1} - R_i|$ is less than

$$2^{-i} p^2 \sqrt{2r_i + \gamma / \log n}$$

and the number of choices for the sequence R_1, R_2, \dots, R_{i+1} is at most

$$n 3 2^{i+1} \log(\alpha'' r'_i + \beta'').$$

Let

$$s_i = \alpha'' r'_i + 2r_i + (\beta'' + \gamma) / \log n.$$

Then the area of $|R_{i+1} - R_i|$ is at most

$$2^{-i} p^2 \sqrt{s_i}$$

and the number of choices for R_1, R_2, \dots, R_{i+1} is at most

$$n 3 2^{i+1} \log(s_i \log n).$$

This proves Lemma 3.2.11 ■.

3.2.6. Probabilistic Part of Proof

We now have all the information needed to bound the discrepancy. Before going into the details needed to do this rigorously, we briefly sketch the proof of the result.

We have that the total number of possibilities for $R_{i+1} - R_i$ is at most

$$n^3 2^{i+1} \log(s_i \log n)$$

and that the area of $|R_{i+1} - R_i|$ is at most

$$2^{-i} p^2 \sqrt{s_i}.$$

We would like to show that the number of points of X in $R_{i+1} - R_i$ is close to its expected value. The standard deviation of the number of points inside a region of area A is $O(\sqrt{A})$, so the standard deviation of the number of points in $R_{i+1} - R_i$ is $\sqrt{\text{Area}(R_{i+1} - R_i)}$, which is $ps_i^{1/4} 2^{-i/2}$. If the distribution is close to normal, the probability of being off more than k standard deviations is e^{-k^2} . If we choose k such that the number of choices for $|R_{i+1} - R_i|$ is less than 2^{k^2} , then the probability that we exceed k standard deviations is $\left(\frac{2}{e}\right)^{k^2}$, which is small for large k . When dealing with the i th approximation, we thus choose k such that $k^2 \approx 2^{i+1} \log(s_i \log n)$. Since the standard deviation is \sqrt{A} , we have that with high probability, the i th approximation adds less than $\sqrt{k^2 A}$ to the area, which is

$$\sqrt{2^{i+1} \log(s_i \log n) 2^{-i} p^2 \sqrt{s_i}} = \sqrt{2} p s_i^{1/4} \sqrt{\log(s_i \log n)}.$$

The total discrepancy is therefore bounded by

$$\sqrt{2} p \sum_{i=1}^m s_i^{1/4} \sqrt{\log(s_i \log n)}.$$

However, $\sum_{i=1}^m s_i = C$. Thus, we need to find the maximum of

$$\sum_{i=1}^m s_i^{1/4} \sqrt{\log(s_i \log n)},$$

given $\sum s_i \leq C$. The function $f(x) = x^{1/4} \sqrt{\log kx}$ is convex. By Jensen's inequality, the maximum occurs when all the s_i are equal. This maximum has a value of $\Theta(\log^{3/4} n)$, giving a discrepancy of $O(\log^{3/4} n)$.

There are several details we have left out in this intuitive discussion. The distribution is not a Gaussian, and is only approximated by a Gaussian in an area close to the mean, so we must make sure we stay in this area. The top $\Theta(\log \log n)$ levels must be taken care of separately, as the variance of their discrepancies is too high for the proof the high probability bound unless we take care of them separately. Finally, we must bound the discrepancies of all the possibilities for $R_{i+1} - R_i$ at once.

Suppose that we have a region of area A . The expected number of points it will contain is A , and the chance of it containing exactly k points is simply $\binom{n}{k} \left(\frac{A}{n}\right)^k \left(\frac{n-A}{n}\right)^{n-k}$. We use this to prove the following lemma:

Lemma 3.2.19: If a region S has area A , and if $\sigma^2 \leq A/2$, then the probability that $\Delta(S) \geq \sigma\sqrt{A}$ is less than $e^{-\sigma^2/3}$. Here, discrepancy, as usual, means $|A - k|$, where k is the number of points in the region.

Proof: The proof is a binomial coefficient manipulation. The chance of a discrepancy of exactly δ is

$$\binom{n}{A+\delta} \left(\frac{A}{n}\right)^{A+\delta} \left(\frac{n-A}{n}\right)^{n-A-\delta}.$$

By Stirling's formula, this is approximately

$$\begin{aligned} & \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(\frac{A}{n}\right)^{A+\delta} \left(\frac{n-A}{n}\right)^{n-A-\delta}}{2\pi\sqrt{(A+\delta)(n-A-\delta)} \left(\frac{A+\delta}{e}\right)^{A+\delta} \left(\frac{n-A-\delta}{e}\right)^{n-A-\delta}} \\ & \leq \left(\frac{A}{A+\delta}\right)^{A+\delta} \left(\frac{n-A}{n-A-\delta}\right)^{n-A-\delta} \\ & = e^{-((A+\delta)\log(1+\frac{\delta}{A}) + (n-A-\delta)\log(1-\frac{\delta}{n-A}))}. \end{aligned}$$

Using the formula $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$, the exponent of the above expression becomes

$$-\left(\delta + \frac{\delta^2}{2A} - \frac{\delta^3}{3A^2} + \dots - \delta + \frac{\delta^2}{2(n-A)} + \frac{\delta^3}{6(n-A)^2} + \dots\right).$$

If $\delta = \sigma\sqrt{A}$, this is at least

$$-\left(\frac{\sigma^2}{2} + \frac{\sigma^2}{2(n-A)} - \frac{\sigma^3}{6\sqrt{A}}\right).$$

If $\sigma^2 \leq A/2$, we have

$$\frac{\sigma^2}{2} - \frac{\sigma^3}{6\sqrt{A}} \leq \frac{\sigma^2}{3},$$

proving the lemma. Note that if we remove the restriction that $\sigma^2 \leq A/2$, the lemma is no longer true. ■

We now need another lemma. This lemma states that if you have a larger region, then the probability of a large discrepancy increases.

Lemma 3.2.20: If there are two regions S_1 and S_2 , and $Area(S_1) \geq Area(S_2)$, then

$$\Pr(\Delta(S_1) \geq k) \geq \frac{1}{3} \Pr(\Delta(S_2) \geq k).$$

Proof: We choose a region $S_3 \subseteq S_1$ of the same area as S_2 . Since S_3 and S_2 are the same area,

$$\Pr(\Delta(S_2) \geq k) = \Pr(\Delta(S_3) \geq k).$$

Now, the probability that the discrepancy of $S_1 - S_3$ has the same sign as that of S_3 is at least $\frac{1}{3}$, so with probability at least $\frac{1}{3}$, $\Delta(S_1) \geq \Delta(S_3)$. This gives the result. ■

Using Lemma 3.2.20, we can see that Lemma 3.2.19 also holds for the discrepancies of signed regions, such as $R_{i+1} - R_i$. A signed region S can be decomposed into two regions R^+ and R^- , with $Area(R^+) + Area(R^-) = Area(R)$ and $\Delta(R^+) + \Delta(R^-) \geq \Delta(R)$. We can then use the above lemmas to bound the discrepancies of R^+ and R^- separately. By Lemma 3.2.20, we can assume that the area of both R^+ and R^- is $Area(|R|)$, since this increases the probability of a large discrepancy. We then use Lemma 3.2.19 to obtain the bound.

Theorem 3.2.21: The total sum of the discrepancies of the regions $R_{i+1} - R_i$, for i satisfying $2 \log \log n \leq i \leq m$, is with probability $1 - n^{-\log^{1/2} n}$ at most $O(p \log^{3/4} n)$.

Proof: Recall the number of choices for R_{i+1} was at most

$$n^3 2^{i+1} \log(s_i \log n) = 2^{2^{i+1}} \log(s_i \log n) + \log^3 n.$$

If $i > 2 \log \log n$, then we can ignore the n^3 term in the number of choices, since

$$2^{i+1} \log(s_i \log n) \geq \log^2 n \log(s_i \log n) \gg 3 \log n,$$

and we can increase s_i to $(1 + \epsilon)s_i$ to absorb the $3 \log n$ term.

We will now introduce the new variable t_i . This variable will essentially be the same as s_i . We require t_i to have a certain minimum value which we will need in some calculations, and we require t_i to be 2^j for some (possibly negative) integer j . This will reduce the number of possibilities for t_i . We will require t_i to have the following properties:

1. $t_i \geq (1 + \epsilon)s_i$
2. $t_i \geq 2 / \log n$.
3. $t_i \geq 1/m$.
4. $t_i = 2^j$ for some (possibly negative) integer j .

It is easy to see that if we always choose the minimum value of t_i satisfying these conditions, then $\sum_i t_i \leq C$ for some new constant C . We thus have that for $i \geq 2 \log \log n$, the number of possible choices for $R_{i+1} - R_i$ is less than $2^{2^{i+1}} \log(t_i \log n)$, and $\text{Area}(|R_{i+1} - R_i|) \leq p^2 \sqrt{t_i} / \Sigma^i$

Now, we claim that with high probability, any possible region $R_{i+1} - R_i$ satisfies

$$\Delta(R_{i+1} - R_i) \leq 3pt_i^{1/4} \sqrt{\log(t_i \log n)}.$$