### Problem Set 3 Solutions

Point totals are in the margin; the maximum total number of points was 57.

1. **Another unbiased estimator for the permanent**

   (a) Denoting the entries of $B$ by $b_{ij}$, we have

   $$X_A = (\det(B))^2 = \left( \sum_{\sigma} \text{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} \right)^2 = \sum_{\sigma} \left( \prod_{i=1}^{n} b_{i\sigma(i)} \right)^2 + \sum_{\sigma \neq \sigma'} \text{sgn}(\sigma \sigma') \left( \prod_{i=1}^{n} b_{i\sigma(i)} \right) \left( \prod_{j=1}^{n} b_{j\sigma'(j)} \right),$$

   where the sums are over all permutations $\sigma, \sigma' \in S_n$. Now note that each diagonal term $\left( \prod_{i} b_{i\sigma(i)} \right)^2 = \prod_{i} b_{i\sigma(i)}^2$ is 1 (with probability 1) if the corresponding diagonal $\prod_{i} a_{i\sigma(i)} = 1$, and zero otherwise. Also, each cross term $\left( \prod_{i} b_{i\sigma(i)} \right) \left( \prod_{j} b_{j\sigma'(j)} \right)$ has expectation zero (because it includes at least one independent factor $b_{ik}$ only once, and $b_{ik}$ is either identically zero or has expectation zero). So, taking expectations, we get

   $$E(X_A) = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma(i)} = \text{per}(A),$$

   and thus $X_A$ is an unbiased estimator.

   (b) Let $A_n$ be the $2n \times 2n$ block-diagonal matrix with $2 \times 2$ all-1s matrices along the diagonal. Clearly $\text{per}(A_n) = 2^n$. Letting $B_n$ denote the corresponding random matrix $B$, we see that $\det(B_n)$ is the product of the determinants of each of its random $2 \times 2$ diagonal blocks; and each of these is zero with probability $\frac{1}{2}$ and ±1 with probability $\frac{1}{2}$. Hence $X_{A_n} = (\det(B_n))^2$ is zero with probability $1 - \frac{1}{2^n}$, and $4^n$ with probability $\frac{1}{2^n}$. Thus in any polynomial number of trials of this estimator, the result returned will be zero with all but exponentially small probability. We can also quantify the variance of the estimator, as measured by the “critical ratio” $\frac{E(A_n^2)}{E(A_n)^2}$, recall that this determines the number of trials needed for a good estimate via Chebyshev’s inequality. This value is $\frac{E(A_n^2)}{E(A_n)^2} = \frac{8^n}{3^n} = 2^n$, which is exponentially large in $n$.

   *Most people gave the above example, and showed that $\text{Var}(A_n)$ is exponential in $n$. However, this isn’t quite enough: since the expectation itself is exponentially large, we could still have large variance even with an efficient estimator. The important quantity is the ratio of the variance to the square of the mean (the “critical ratio”). An alternative justification is to simply observe that the algorithm outputs 0 with all but exponentially small probability.*

   (c) In similar fashion to part (a), the numerator can be written as

   $$E(X_A^2) = E(\det(B)^4) = E\left( \left( \sum_{\sigma} \text{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} \right)^4 \right).$$

   Reasoning again as in part (a), we see that the only terms that survive in this quartic expansion are those in which each $b_{i\sigma(i)}$ appears an even number (2 or 4) of times (else the expectation of the term is zero). Each such term has expectation 1. Now we can associate each such term to a pair of perfect matchings $(M, M')$ as follows. Note that the term itself consists of the product of four diagonals $\prod_{i} b_{i\sigma(i)}$, each of which corresponds to a perfect matching $\sigma$ in $G_A$. Call these four matchings $\sigma_1, \sigma_2, \sigma_3, \sigma_4$. The constraint that each edge appears an even number of times means that the union of the four matchings...
is in fact equal to the union of just two matchings (with each edge coinciding with two or four edges of the $\sigma_i$). We call this pair of matchings $\langle M, M' \rangle$; and, since the pair is ordered, we may assume w.l.o.g. that $M$ coincides with $\sigma_1$. Now we claim that the number of quadruples $\langle \sigma_1, \sigma_2, \sigma_3, \sigma_4 \rangle$ that give rise to a given pair $\langle M, M' \rangle$ in this way is exactly $3^{c(M,M')}$. To see this, consider each cycle in the union $\langle M, M' \rangle$. We already know that $\sigma_1$ coincides with $M$ on this cycle. Then we may choose any one of $\sigma_2, \sigma_3, \sigma_4$ to coincide with $\sigma_1$, with the other two matchings coinciding on the complementary edges of the cycle. So we have three choices per cycle, and these choices are made independently for each cycle, giving a total contribution of $3^{c(M,M')}$ for each pair $\langle M, M' \rangle$. (Edges shared by $M, M'$, which are not counted as cycles, can occur in only one way, namely when all four of the $\sigma_i$ coincide on that edge.)

To conclude, note from part (a) that the denominator in the critical ratio, $E(X_A)^2$, is just $\text{per}(A)^2$, which is equal to the number of ordered pairs of perfect matchings $\langle M, M' \rangle$ in $G_A$. Hence we may view the quotient in the critical ratio as the expectation of the quantity in the numerator, namely $3^{c(M,M')}$, when the pair $\langle M, M' \rangle$ is chosen u.a.r. Hence the critical ratio is indeed equal to $\gamma(A)$.

(d) Using Chebyshev and part 2 of Lemma 13.10 as at the bottom of page 13-5, we obtain

$$\Pr_{A_{n,m}}[\text{per}(A)^2 \leq \frac{9}{16}(E_{A_{n,m}}(\text{per}(A)))^2] = \Pr_{A_{n,m}}[\text{per}(A) \leq \frac{3}{4}E_{A_{n,m}}(\text{per}(A))] = O\left(\frac{n^3}{m^2}\right).$$

But another (direct) application of part 2 of Lemma 13.10 yields

$$\frac{1}{2}(E_{A_{n,m}}(\text{per}(A))^2) \leq \frac{1}{2}(1 + O(\frac{n^3}{m^2}))(E_{A_{n,m}}(\text{per}(A)))^2 \leq \frac{9}{16}(E_{A_{n,m}}(\text{per}(A)))^2,$$

for sufficiently large $n$. Putting these two facts together gives the desired claim.

(e) The given fact about the expected value of $3^{c(M,M')} \langle M, M' \rangle$ can be written as

$$\frac{1}{|\Omega|} \sum_{A \in A_{n,m}} \text{per}(A)^2 \gamma(A) \leq Cn^2. \quad (1)$$

Now assume for the sake of contradiction that, for some $\varepsilon > 0$, we have

$$\Pr_{A_{n,m}}[\gamma(A) \geq n^2\omega(n)] \geq \varepsilon \quad \text{for infinitely many } n. \quad (2)$$

Combining this with part (d), we deduce that at least a fraction $\varepsilon - O(\frac{n^3}{m^2}) \geq \frac{\varepsilon}{2}$ of all matrices $A \in A_{n,m}$ simultaneously satisfy $\gamma(A) \geq n^2\omega(n)$ and $\text{per}(A)^2 \geq \frac{1}{4}E_{A_{n,m}}(\text{per}(A)^2)$. But this implies that

$$\frac{1}{|\Omega|} \sum_{A \in A_{n,m}} \text{per}(A)^2 \gamma(A) \geq \frac{1}{|\Omega|} \times \frac{\varepsilon}{2} |A_{n,m}| \times \frac{1}{2}E_{A_{n,m}}(\text{per}(A)^2) \times n^2\omega(n) = \frac{\varepsilon}{4}n^2\omega(n),$$

which contradicts (1). Thus our assumption in (2) must be false, which implies that $\Pr_{A_{n,m}}[\gamma(A) \geq n^2\omega(n)] \to 0$ as desired.

(f) Note that each trial of the estimator $X_A$ takes polynomial time to compute as it just involves a single determinant computation. By the unbiased estimator theorem (Theorem 11.1 in Lecture 11), it therefore suffices to show that the critical ratio $\frac{E(X_A^2)}{E(X_A)^2}$ is polynomially bounded with high probability over $A \in A_n$, since the number of trials needed is proportional to this ratio. By part (c) this ratio is precisely $\gamma(A)$, and by part (e) the ratio is bounded by $n^2\omega(n)$ with high probability over $A \in A_{n,m}$ when $\frac{m^2}{n^2} \to \infty$. But now we can follow the same argument as in the proof of Lemma 13.8 to translate this statement to $A_n$. Since the number $m$ of 1’s in a random $A \in A_n$ is tightly concentrated about its mean $\frac{n^2}{2}$, we have that $\frac{m^2}{n^2} \to \infty$ w.h.p. Hence we have also $\Pr_{A_n}[\gamma(A) \leq n^2\omega(n)] \to 1$ as $n \to \infty$. Many people forgot the final step above converting from the distribution $A_{n,m}$ to $A_n$.  

2pts
2. Chernoff for Poisson

(a) We start with the observation that, for a Poisson r.v. \( X \) with parameter \( \mu \) and any \( t > 0 \), we have

\[
E[e^{tX}] = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{tk} = e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu e^t)^k}{k!} = \exp(\mu(e^t - 1)).
\]

Now, following the standard argument for such bounds (see Lecture 14), we have for any \( t > 0 \)

\[
\Pr[X \geq \mu + \lambda] = \Pr[e^{tX} \geq e^{(\mu+\lambda)t}] \leq e^{-(\mu+\lambda)t} E[e^{tX}] = e^{-(\mu+\lambda)t} \exp(\mu(e^t - 1)) = \exp(\mu(e^t - 1) - (\mu + \lambda)t).
\]

By elementary calculus, the value of \( t \) that minimizes this bound is given by \( e^t = \frac{\mu + \lambda}{\mu} \). Plugging this in and tidying up gives the desired bound:

\[
\Pr[X \geq \mu + \lambda] \leq \exp \left\{ -[\mu + \lambda] \ln \left( \frac{\mu + \lambda}{\mu} \right) - \lambda \right\}.
\]

A symmetrical argument for the lower tail yields

\[
\Pr[X \leq \mu - \lambda] \leq \exp \left\{ -[\mu - \lambda] \ln \left( \frac{\mu - \lambda}{\mu} \right) + \lambda \right\}.
\]

(b) Setting \( \lambda = \beta \mu \) in the above bound gives

\[
\Pr[X \geq (1 + \beta)\mu] \leq \exp \left\{ -\mu[(1 + \beta) \ln(1 + \beta) - \beta] \right\},
\]

which is exactly the same as Angluin’s bound for the binomial distribution (Corollary 14.3). The same substitution for the lower tail again recovers Angluin’s bound:

\[
\Pr[X \leq (1 - \beta)\mu] \leq \exp \left\{ -\mu[(1 - \beta) \ln(1 - \beta) + \beta] \right\}.
\]

3. Random geometric graphs

(a) Following the hint, partition the unit square into small squares of area \( \frac{\log n}{n} \). Let \( S \) be any such square. Then

\[
\Pr[S \text{ contains no points}] = (1 - \text{area}(S))^n = (1 - \frac{\log n}{n})^n \leq e^{-\log n} = n^{-1}.
\]

Thus taking a union bound over all \( \frac{n}{\log n} \) choices of \( S \), we see that \( \Pr[\text{some } S \text{ contains no points}] \to 0 \) as \( n \to \infty \).

Now let \( D \) denote the disc of radius \( \sqrt{\frac{10 \log n}{n}} \) centered at some point. W.l.o.g. we may assume that \( D \) is contained entirely within the unit square; otherwise the argument below only gets better. The expected number of points in \( D \) is \( \mu := n \times \text{area}(D) = 10\pi \log n \). Thus using Angluin’s version of the Chernoff bound (upper tail with \( \beta = 1 \)), we get

\[
\Pr[D \text{ contains more than } 2\mu \text{ points}] \leq \exp(-\frac{\mu}{3}) = \exp(-\frac{10\pi}{3} \log n) = n^{-(1+\delta)},
\]

for some constant \( \delta > 0 \). Taking a union bound over all \( n \) points ensures that

\[
\Pr[\text{some } D \text{ contains more than } 2\mu \text{ points}] \to 0 \quad \text{as } n \to \infty.
\]

Now take \( c > 20\pi \), and form the graph \( G \) in which every point is connected to its \( c \log n \) closest neighbors. From the above arguments we have that, with probability \( 1 - o(1) \), every square contains a point, and every point is connected to points in all its immediately neighboring squares. But these conditions are clearly enough to ensure that \( G \) is connected.

Some people forgot to take a union bound over all \( n \) points in order to ensure that every point is connected to the points in each neighboring square.
(b) Suppose that the point set contains a “bad” system of three concentric discs, as defined in the question. Note first that conditions (i) and (ii) imply that the \( k \) nearest neighbors of all points in \( D_1 \) also lie in \( D_1 \) and condition (ii) says that there are no points in \( D_3 \setminus D_1 \). We now argue that conditions (ii) and (iii) imply that the \( k \) nearest neighbors of any point outside \( D_3 \) all lie outside \( D_3 \). To see this, for any point \( p \) outside \( D_3 \), let \( r_p \) denote the distance to the boundary of \( D_1 \) (thus \( r_p \) is a lower bound on the distance to the nearest point in \( D_1 \)). Consider the disc of radius 1.5\( r \) (as defined in (iii)) centered at a point on \( D_3 \) that is closest to the line connecting \( p \) with the boundary of \( D_1 \). By the triangle inequality, the distance from \( p \) to this disc’s center will be at most \( r_p - 2r + .01r \), and by condition (iii) and the triangle inequality, there will be at least \( k \) points within distance \( r_p - 2r + .01r + 1.5r = r_p - .49r \) from the point \( p \), and thus \( p \) will not be connected to any point inside \( D_1 \), as desired.

(c) Probabilities of occupancy events \( E \) in the \( n \)-point model and the PPP of intensity \( n \) are related by

\[
Pr[E] = Pr_{Po(n)}[E | Po(n) = n] \leq \frac{Pr_{Po(n)}[E]}{Pr_{Po(n)}[Po(n) = n]}.
\]

(This reflects the fact that the only dependencies between point positions in the \( n \)-point model arise from the fact that the number of points is fixed.) Now from the definition of the PPP, the probability that the number of points generated is exactly \( n \) is \( Pr_{Po(n)}[Po(n) = n] \geq \frac{1}{e \sqrt{n}} \). Plugging this into equation (3) gives

\[
Pr[E] \leq e \sqrt{n} \times Pr_{Po(n)}[E] \leq e \sqrt{n} \times \delta(n) \to 0,
\]
as required.

(d) Now let’s consider an arbitrary system of three concentric discs as above. We analyze the probability of each of the events (i), (ii), (iii) under the PPP model.

(i) The expected number of points in disc \( D_1 \) is \( \mu := n \times \text{area}(D) = k + 1 \), so we are looking for the probability that a Poisson r.v. is at least equal to its mean. This is at least 1/4. (Actually it is close to 1/2.)

(ii) The expected number of points in \( D_3 \setminus D_1 \) is \( n \times \text{area}(D_3 \setminus D_1) = 8 \pi r^2 n = 8(k+1) \). Thus the probability that no points fall in this area is \( \exp(-8(k+1)) = \exp(-8(c_2 \log n + 1)) \). If we choose \( c_2 > (1-\epsilon)/8 \) for some small \( \epsilon > 0 \), this probability is greater than \( n^{-1 + \epsilon} \).

(iii) The expected number of points lying in each circle of radius 1.5\( r \), excluding the portion within \( D_3 \), is at least \( \frac{\pi(1.5r)^2}{2} = 1.125(k+1) \), and thus the probability that a given circle contains fewer than \( k + 1 \) points, by the Chernoff bound for Poissons in Q2 above, is \( \exp(-c_3 k) \), for some constant \( c_3 \). Applying the union bound to the \( \frac{\pi 6r}{30r} \leq 2000 \) such circles on the boundary of \( D_3 \) yields that this condition is satisfied with probability at least \( 1 - 2000 \exp(-c_3 k) > .99 \), for sufficiently large \( n \).

Putting together the above three events, and noting that they are independent since they refer to disjoint areas, we get that

\[
Pr[\text{set of discs is bad}] \geq (1/4)(.99)n^{-1 + \epsilon} \geq c' n^{-1 + \epsilon}
\]

for some constant \( c' \). Finally, note that we can pack a total of \( c'' \frac{n}{\log n} \) disjoint systems of three discs into our unit square, for some absolute constant \( c'' \). Each of these is bad independently with probability at least \( c' n^{-1 + \epsilon} \), so the probability that no bad set of discs exists in the PPP model is at most

\[
(1 - c' n^{-1 + \epsilon})^{c'' n/\log n} \leq \exp(-c n^\epsilon/\log n).
\]

(e) Since \( \exp(-c n^\epsilon/\log n) \ll 1/\sqrt{n} \), we conclude from parts (c) and (d) that, in the original \( n \)-point model, the probability that no bad set of discs exists tends to 0 as \( n \to \infty \). By part (b), this implies that the probability that the graph is connected tends to 0 as \( n \to \infty \).
4. Codes in space

(a) Consider one pair of random strings of length $\ell$. The probability that they agree in more than $\epsilon \ell$ positions is the probability that the number of successes $X$ in $\ell$ independent trials, each with success probability $\frac{1}{a}$, exceeds $\epsilon \ell$. Using the Chernoff bound $\Pr[X > \mu + \lambda] \leq \exp(-2\lambda^2/n)$ from Corollary 14.2 of Lecture 14, with $\mu = \frac{\ell}{a}$ and $\lambda = (\epsilon - \frac{1}{a})\ell$, we get

$$\Pr[X > \epsilon \ell] \leq \exp(-2(\epsilon - \frac{1}{a})^2 \ell) \leq \exp(-2(\epsilon - \frac{1}{a})^2 C \ln m) = m^{-2(\epsilon/a)^2 C}.$$ 

By choosing $C > (\epsilon - 1/a)^{-2}$ we can make this $o(m^{-2})$, so taking a union bound over all $\binom{m}{2}$ pairs of strings ensures that they form an $(m, \ell, \epsilon)$-code with probability $\rightarrow 1$.

(b) Using the same analysis as above, but the stronger Chernoff bound quoted in the hint, with $\beta = \frac{1}{\mu} = 3$ and $ae - 1$, we get

$$\Pr[X > \epsilon \ell] \leq \exp \left\{ -\frac{\ell}{a} (ae \ln(ae) - ae + 1) \right\} \leq \exp \left\{ -\ell (\epsilon a\ln(ae) - \epsilon) \right\}.$$ 

Now for any fixed $\epsilon > 0$, by choosing $a$ large enough, we can make this probability less than $\exp(-K\ell)$ for any desired $K$. So if $\ell = \delta \ln m$ for fixed $\delta > 0$, by arranging for $K > \frac{2}{a}$ we can still make the above $o(m^{-2})$, and then apply a union bound over pairs of strings as in part (a). (Specifically, we just need that $a > \epsilon^{-1} \exp(1 + 2/(\epsilon \delta))$.)

(c) Any embedding of a pair of strings in $\mathbb{Z}^3$ defines a “set of adjacencies” consisting of those pairs of symbols (from different strings, or non-consecutive positions on the same string) that are adjacent in the embedding. We will bound the number of possible sets of adjacencies, and the probability that any given set has a large score, then use a union bound.

To bound the number of possible sets of adjacencies for a pair of strings, note that each one can be realized by embedding both strings in a cube of side length $2\ell$. Within such a cube, the number of embeddings of the two strings is (very crudely) at most $((2\ell)^3(6\ell-1))^2$, where the first factor counts the starting points and the second the number of walks in $\mathbb{Z}^3$ of length $\ell - 1$, and we pessimistically ignore the constraint that the two strings be non-overlapping. Thus the number of possible sets of adjacencies is at most $\exp(\alpha \ell)$ for some universal constant $\alpha$.

Now fix a particular set of adjacencies. The maximum number of possible pairs in the set is at most $5\ell$ (since each of the $2\ell$ symbols on the strings can be adjacent to at most 5 others, and each adjacency gets counted twice). For each pair, we get a score of 1 iff the two symbols in it are assigned the same value. Thus, the probability of getting a score of more than $\epsilon \ell$ is bounded by the probability of more than $\epsilon \ell$ successes in $5\ell$ trials with success probability $\frac{1}{a}$. However, the events that different pairs score are not necessarily independent. (E.g., consider a cycle $(p_1, p_2, p_3, p_4)$ of length 4 in $\mathbb{Z}^3$, and suppose the pairs $(p_1, p_2)$, $(p_2, p_3)$, $(p_3, p_4)$ and $(p_4, p_1)$ are adjacencies; if any three of these score, then so must the fourth.) To get around this, we can partition the adjacencies into three disjoint sets according to the direction (in $\mathbb{Z}^3$) of the adjacency. Now it’s easy to see that the scoring events within each set of adjacencies are independent. Moreover, it is sufficient to bound the probability of a score of more than $\frac{5\ell}{2}$ within each such set, and multiply by 3. Using the same Chernoff bound as in part (b), but now with $\mu = \frac{5\ell}{15}$ and $\beta = \frac{ae}{15} - 1$, we see that this probability is bounded by

$$3 \exp \left\{ -\frac{5\ell}{a} \left( \frac{ae}{15} \ln(\frac{ae}{15}) - \frac{ae}{15} + 1 \right) \right\} \leq 3 \exp \left\{ -\ell \left( \frac{\epsilon}{3} \ln(\frac{ae}{15}) - \frac{\epsilon}{3} \right) \right\}.$$ 

Now, as in part (b), we can choose $a$ large enough so that this probability is less than $\exp(-K\ell)$ for any desired $K$. Thus, taking the union bound over sets of adjacencies, the probability of two strings having a score of more than $\epsilon \ell$ in any embedding is at most $\exp(\alpha \ell) \exp(-K\ell) = o(m^{-2})$ if we set $\ell = \delta \ln m$ and choose $a$ (and thus $K$) large enough. Again, a union bound over pairs of strings finishes the job.
5. More on the power of two choices

Following the hint, we will define a decreasing sequence of values $\alpha_i$ (to be determined shortly) and events $\mathcal{E}_i = \{ \text{after } (1 - \frac{1}{2r})n \text{ balls have been thrown, the number of bins with load } \geq i \text{ is at least } \alpha_i \}$.

Our aim is to bound the probability $\Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i]$. So assume that $\mathcal{E}_i$ holds, and consider the placement of balls $(1 - \frac{1}{2r})n + 1$ through $(1 - \frac{1}{2r+1})n$ (of which there are $\frac{n}{2r+1}$). For any such ball, if both of its choices are bins with load at least $i$, and at least one of the two has load exactly $i$, then a new bin with load $i + 1$ will be created.

Assuming that we still have fewer than $\alpha_{i+1}$ bins of load at least $i + 1$ when the ball is thrown, the probability that the ball makes such choices is at least

$$\left( \frac{\alpha_i}{n} \right) \left[ \left( \frac{\alpha_i}{n} \right) - \left( \frac{\alpha_{i+1}}{n} \right) \right] \geq \frac{1}{2} \left( \frac{\alpha_i}{n} \right)^2,$$

where we have assumed that $\alpha_i \geq \frac{1}{2} \alpha_{i+1}$. (This is for algebraic convenience only; in fact $\alpha_i$ will decrease quite a bit faster than this; see below.) To justify (*), note that the number of bins with load exactly $i$ remains at least $(\alpha_i - \alpha_{i+1})$ throughout, as long as the number of bins with load at least $i + 1$ remains below $\alpha_{i+1}$; and the the number of bins with load at least $i$ is certainly at least $\alpha_i$.

Noting that if the number of bins with load at least $i + 1$ reaches $\alpha_{i+1}$ then event $\mathcal{E}_{i+1}$ certainly holds, we see from (*) that the desired probability $\Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i]$ is bounded above by the probability that $\text{Bin}(\frac{n}{2r+1}, \frac{1}{2} \left( \frac{\alpha_{i+1}}{n} \right)^2)$ is less than $\alpha_{i+1}$. The expectation of the above random variable is $\mu_i = \frac{1}{2(2r+1)} \alpha_{i+1}^2$. To get a small tail probability, this suggests that we should take (say) $\alpha_{i+1} = \frac{1}{2} \mu_i$, or equivalently, $\alpha_{i+1} = \frac{1}{2r+3} \frac{\alpha_i^2}{n}$. Of course, $\alpha_0 = n$.

With this choice of $\alpha_i$, a Chernoff bound (Angluin’s version, Corollary 14.3) with $\beta = \frac{1}{2}$ tells us that

$$\Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i] \leq \exp(-\mu_i/8) = O(n^{-1})$$

provided $\mu_i \geq 8 \ln n$. Let $i^*$ be the largest integer $i$ for which this still holds. By unwinding the recurrence $\frac{\alpha_{i+1}}{n} = \frac{1}{2r+3} \left( \frac{\alpha_i}{n} \right)^2$, we see that $\alpha_i = n 2^{-\Theta(2^r)}$ and hence $i^* = \frac{\ln \ln n}{\ln 2} - O(1)$. Finally we have

$$\Pr[\mathcal{E}_{i^*}] \geq \Pr[\mathcal{E}_0] \times \prod_{i=0}^{i^*-1} \Pr[\mathcal{E}_{i+1} | \mathcal{E}_i] \geq \left( 1 - \frac{1}{n} \right)^{O(\ln \ln n)} = 1 - o(1).$$

But since $\alpha_{i^*} \geq 1$, $\mathcal{E}_{i^*}$ implies that the maximum load is at least $i^*$ and we are done.

Some people failed to realize that a new bin with load $i + 1$ will only be created if at least one of the two possible choices has exactly $i$ balls.