Probability and Computing, 2nd Edition

Solutions to Chapter 5: Balls, Bins, and Random Graphs

Hahndeul Kim

July 2025

5.1

As $(1 + 1/n)^n$ increases, we find the smallest n to reach the threshold.
$(1 + 1/n)^n$ first reaches $0.99e$ at $n = 50$, and $0.999999e$ at $n = 499982$.
Since $(1 - 1/n)^n$ also increases, we solve in a similar way. $(1 - 1/n)^n$ first reaches $0.99/e$ at $n = 51$ and $0.999999/e$ at $n = 499991$.

5.2

Recall the formula used in the birthday paradox: If there are N possibilities, then we solve for the smallest n that satisfies $\prod_{i=1}^{n-1}(1 - \frac{i}{N}) \approx \prod_{i=1}^{n-1} e^{-i/n} = e^{-(n-1)n/2N} < 1/2$. Note that we omitted the final approximation to derive exact numerical answers.

Regardless of whether the number of Social Security number digits is 9 or 13, using the last four digits gives $N = 10000$ and this gives $n = 119$.

In the case where the number of digits is 9 ($N = 10^9$), we get $n = 37234$.

In the case where the number of digits is 13 ($N = 10^{13}$), we get $n = 3723298$.

5.3

Let the number of balls thrown be m. Then the desired probability is $\prod_{i=0}^{m-1}(1 - \frac{i}{n})$.

We first determine c_1. $m = c_1\sqrt{n}$ should satisfy $\prod_{i=0}^{m-1}(1 - \frac{i}{n}) \leq \prod_{i=0}^{m-1} e^{-i/n} = e^{-(m-1)m/2n} \leq e^{-1}$. Since $(m-1)m = c_1^2 n - c_1\sqrt{n} \geq 2n$, $(c_1^2 - 2)\sqrt{n} \geq c_1$.

Therefore, we choose c_1 that is greater than or equal to $\frac{1}{2}\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{1}{n} + 8}\right)$.

Now we determine c_2. To use the given hint, assume that $2m < n$.

$\prod_{i=0}^{m-1}(1 - \frac{i}{n}) \geq \prod_{i=0}^{m-1} \exp(-\frac{i}{n} - \frac{i^2}{n^2}) = \exp(-\frac{m(m-1)}{2n} - \frac{(m-1)m(2m-1)}{6n^2})$

$= \exp(-\frac{m(m-1)}{2n}(1 + \frac{2m-1}{3n})) \geq \exp(-\frac{m^2}{2n}(1 + \frac{2m}{3n})) \geq \frac{1}{2}$ should be satisfied for $m = c_2\sqrt{n}$. This is equivalent to satisfying $\frac{c_2^2}{2}(1 + \frac{2c_2}{3\sqrt{n}}) \leq \ln 2$.

Since n is sufficiently large, choosing $c_2 = \sqrt{2\ln 2 - \frac{1}{\ln n}}$ yields the desired result.

5.4

Let event A indicate that there exist two or more people who share a birthday, and event B indicate that exactly two people share a birthday. Then our desired probability would be $\Pr(A - B) = \Pr(A) - \Pr(B)$ since $B \subset A$.

We first determine $\Pr(A)$, which is easy: $\Pr(A) = 1 - \prod_{i=1}^{100} \frac{366-i}{365}$.

We now determine $\Pr(B)$. If there are i shared birthdays in which each day is

shared by exactly two people, then the number of possible permutations would be the product of the following terms:

$\binom{365}{i}$ ways to choose i shared days, $\binom{100}{2i}$ ways to choose $2i$ people to share birthdays, $\prod_{j=1}^{i} \binom{2j}{2}$ ways to distribute i birthdays to $2i$ people and $\prod_{j=1}^{100-2i} (366 - i - j)$ ways to distribute unique birthdays to the rest.

Thus, $\Pr(B) = \sum_{i=0}^{50} \frac{365!100!}{i!(100-2i)!(265+i)!2^i} \times \frac{1}{365^{100}}$.

Therefore, we can determine our desired probability $\Pr(A-B) = \Pr(A) - \Pr(B)$.

5.5

Let $X \sim Poisson(\lambda)$. Then $M_X(t) = \mathbf{E}[e^{tX}] = e^{\lambda(e^t - 1)}$ holds. By computing the second derivative of $M_X(t)$ with respect to t and plugging $t = 0$ in, we get $\mathbf{E}[X^2] = \lambda + \lambda^2$. Thus, $\mathbf{Var}[X] = \lambda$ follows.

5.6

We first show that $Y \sim Poisson(\mu p)$.

$\Pr(Y = k) = \sum_{i=k}^{\infty} \Pr(X = i)\binom{i}{k}p^k(1-p)^{i-k} = \sum_{i=k}^{\infty} \frac{e^{-\mu}\mu^i}{i!} \frac{i!}{k!(i-k)!}p^k(1-p)^{i-k}$

$= \frac{e^{-\mu}(p\mu)^k}{k!} \sum_{i=k}^{\infty} \frac{(\mu(1-p))^{i-k}}{(i-k)!} = \frac{e^{-\mu}(p\mu)^k}{k!}e^{\mu(1-p)} = \frac{e^{-\mu p}(\mu p)^k}{k!}$.

We can also similarly show that $Z \sim Poisson(\mu(1-p))$.

Now we show that $\Pr(Y = i, Z = j) = \Pr(Y = i)\Pr(Z = j)$. Note that $X = Y + Z$ by definition. This allows us to write $\Pr(Y = i, Z = j)$ as $\Pr(Y = i, X = i + j) = \Pr(X = i + j)\binom{i+j}{i}p^i(1-p)^j = \frac{e^{-\mu}\mu^{i+j}}{i!j!}p^i(1-p)^j$.

Since $\Pr(Y = i)\Pr(Z = j) = \frac{e^{-p\mu}(p\mu)^i}{i!}\frac{e^{-(1-p)\mu}((1-p)\mu)^j}{j!} = \frac{e^{-\mu}\mu^{i+j}p^i(1-p)^j}{i!j!} = \Pr(Y = i, X = i + j)$, $Y \perp\!\!\!\perp Z$. $\blacksquare$

5.7

We first prove that $\ln(1 + x) \leq x$, which is equivalent to $1 + x \leq e^x$.

Since $\ln(1 + x) - x = -\frac{x^2}{2} + \frac{x^3}{3} - \cdots$, this can be seen as an alternating series as $\frac{x^n}{n}$ is monotonically decreasing in $|x| \leq 1$. We can apply rearrangements to the alternating series as $\ln(1 + x) - x = -\sum_{n=1}^{\infty} \left(\frac{x^{2n}}{2n} - \frac{x^{2n+1}}{2n+1}\right)$, since the Taylor expansion of $\ln(1 + x)$ is absolutely convergent (to $e^x - 1$). The rearrangement gives $\ln(1 + x) - x \leq 0$, which is the desired result.

We now prove $x + \ln(1 - x^2) \leq \ln(1 + x)$, which is equivalent to $e^x(1 - x^2) \leq 1 + x$. Since $\ln(1 - x^2) = \ln(1 + x) + \ln(1 - x)$, $x + \ln(1 - x^2) \leq \ln(1 + x)$ is reduced to $\ln(1 - x) \leq -x$. At $|x| \leq 1$, this is equivalent to $\ln(1 + x) \leq x$, which we have previously proved. $\blacksquare$

5.8

(a) Since the ball is equally likely to fall in one of the three bins, the desired probability is 1/3.

(b) Since the bin 2 did not receive balls, we can simply think of this as throwing balls n into $n - 1$ bins. The conditional expectation would be $n/(n-1)$.

(c) Note that the probability that bin 1 receives more balls than bin 2 is the same as that of bin 2 receiving more balls than bin 1. Thus, we first compute the probability that two bins receive the same number of balls, which is
$P = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k}\binom{2k}{k}(\frac{1}{n})^{2k}(1-\frac{2}{n})^{n-2k}$. The desired probability would be $(1-P)/2$.

5.9

In the given condition, the expected number of elements in a single bucket is at most a. Since $a = O(1)$, sorting all buckets can still be done in linear time.

5.10

(a) By the Poisson approximation, the probability p is bounded as $p \le e\sqrt{n}(\frac{1}{e})^n$.

(b) $\frac{n!}{n^n}$.

(c) Since $\Pr(Z = n) = \frac{e^{-n}n^n}{n!}$ when $Z \sim Poisson(n)$, $\frac{n!}{n^n} \times \frac{e^{-n}n^n}{n!} = e^{-n}$ shows the claim. Theorem 5.6 states that the distribution $(Y_1, ..., Y_n)$ constrained on $\sum_{i=1}^{n} Y_i = n$ is equivalent to the balls and bins model. Note that each Y_i follows $Poisson(1)$ and each X_i denotes the load of the ith bin in the balls and bins model. Then using theorem 5.6, $(1/e)^n/(\frac{e^{-n}n^n}{n!}) = \frac{\Pr(\forall i, Y_i=1)}{\Pr(\sum_i Y_i=n)} = \Pr(\forall i, Y_i = 1 | \sum_i Y_i = n) = \Pr(\forall i, X_i = 1) = \frac{n!}{n^n}$.

5.11

Let X_i be the indicator variables that are 1 if there is a k-gap starting at i.
Let $X = \sum_{i=0}^{n-k} X_i$.

(a) The expected number of k-gaps would be $\sum_{i=0}^{n-k} \mathbf{E}[X_i] = (n - k + 1)(1 - \frac{k}{n})^m$.

(b) First, we assume that the bin loads follow the Poisson distribution to derive the Poisson-approximated Chernoff bound. We divide $\{X_i\}$ into k subsets, so that all indicator variables in the same subset are independent. First, we derive the Chernoff bound for $Y_0 = \sum_{i \ge 0} X_{ik}$ where $0 < \delta < 1$.

$\Pr(Y_0 \ge (1+\delta)\mathbf{E}[Y_0]) \le \exp(-\frac{1}{3}\frac{\mathbf{E}[X]}{k}\delta^2)$, and $\Pr(Y_0 \le (1-\delta)\mathbf{E}[Y_0]) \le \exp(-\frac{1}{2}\frac{\mathbf{E}[X]}{k}\delta^2)$ holds. Therefore, $\Pr(|Y_0 - \mathbf{E}[Y_0]| \ge \delta\mathbf{E}[Y_0]) \le 2\exp(-\frac{1}{3}\frac{\mathbf{E}[X]}{k}\delta^2)$.

4

With the union bound for all Y_i, we get $\Pr(|X - \mathbf{E}[X]| \geq \delta \mathbf{E}[X]) \leq 2k \exp(-\frac{1}{3} \frac{\mathbf{E}[X]}{k} \delta^2)$. Since we have used the Poisson approximation to compute the Chernoff bound, the computed upper bound should be multiplied by $e\sqrt{m}$.

5.12

Let X_i be the indicator variables that are 1 if a ball landed in bin i by itself.

(a) The expected number of balls to be *served* in this round would be $\sum\limits_{i=1}^{n} \mathbf{E}[X_i] = \sum\limits_{i=1}^{n} b \times \frac{1}{n}(1 - \frac{1}{n})^{b-1} = b(1 - \frac{1}{n})^{b-1}$. Therefore, the expected number of balls at the start of the next round would be $b(1 - (1 - \frac{1}{n})^{b-1})$.

(b) Note that if $n = 1$, then the number of rounds required would be trivially 1. Therefore, we only consider the cases where $n \geq 2$.

Since $x_{j+1} = x_j(1 - (1 - \frac{1}{n})^{x_j - 1}) \leq x_j(1 - (1 - \frac{x_j - 1}{n})) = x_j \frac{x_j - 1}{n} \leq \frac{x_j^2}{n}$, the inequality given in the hint is true.

With $x_1 = n(1 - (1 - \frac{1}{n})^{n-1})$, cascading the inequality yields $x_k \leq n(\frac{x_1}{n})^{2^{k-1}} = n(1 - (1 - \frac{1}{n})^{n-1})^{2^{k-1}} \leq n(1 - (1 - \frac{1}{n})^n)^{2^{k-1}}$.

Now, let k^* be the minimum k that satisfies $n(1 - (1 - \frac{1}{n})^n)^{2^{k-1}} \leq 1$, so that the operation ends after no more than $k^* + 1$ rounds. Since $1 - (1 - \frac{1}{n})^n \leq \frac{3}{4}$ at $n \geq 2$, we calculate the minimum k that satisfies $n(\frac{3}{4})^{2^{k-1}} \leq 1$.

Taking log on both sides gives $\ln n + 2^{k^*-1} \ln \frac{3}{4} \leq 0$, which is equivalent to $2^{k^*-1} \geq \frac{\ln n}{\ln \frac{4}{3}}$. Therefore, we get $k^* - 1 \geq \ln \frac{\ln n}{\ln \frac{4}{3}}$, which shows that $k^* = O(\log \log n)$.

5.13