(5.4-2)

Suppose that balls are tossed into b bins. Each toss is independent, and each ball is equally likely to end up in any bin. What is the expected number of ball tosses before at least one of the bins contains two balls?

When a tossed ball lands in an occupied bin, we call this a collision. Let X be "the number of throws prior to the first collision." Then

$$\mathrm{E}[X] = \sum_{k=0}^{\infty} k \Pr[X = k] = \sum_{k=1}^{\infty} \Pr[X \geq k] = \sum_{k=1}^{b+1} \Pr[X \geq k] \; .$$

At this point, the analysis closely follows that for the birthday paradox (on CLRS p. 107), where

$$\begin{aligned}
\Pr[X \geq k] &= \Pr[X \geq k \mid X \geq k-1] \Pr[X \geq k-1] \\
&= \frac{b-k+1}{b} \Pr[X \geq k-1] \\
&\vdots \\
&= \left(\frac{b-k+1}{b}\right) \cdot \left(\frac{b-k+2}{b}\right) \cdots \left(\frac{b-1}{b}\right) \cdot \frac{b}{b} \\
&= \frac{b!}{b^k (b-k)!} \; .
\end{aligned}$$

So we have that

$$\mathrm{E}[X] = \sum_{k=1}^{b+1} \Pr[X \geq k] = \sum_{k=1}^{b+1} \frac{b!}{b^k (b-k)!} \; .$$

One way to bound this is to rearrange terms into the Taylor series expansion of the exponential function and apply Stirling's approximation:

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=1}^{b+1} \frac{b!}{b^k (b-k)!} = \sum_{k=1}^{b+1} \frac{b^{b-k}}{b^b} \frac{b!}{(b-k)!} \\
&= \frac{b!}{b^b} \sum_{k=1}^{b+1} \frac{b^{b-k}}{(b-k)!} = \frac{b!}{b^b} \sum_{k=0}^{b} \frac{b^k}{k!} \\
&< \frac{b!}{b^b} \sum_{k=0}^{\infty} \frac{b^k}{k!} = \frac{b!}{b^b} e^b \\
&\approx \frac{\sqrt{2\pi b}\left(\frac{b}{e}\right)^b}{b^b} e^b = \sqrt{2\pi b}
\end{aligned}$$

An alternate, but wrong, solution:
Note that the following is *not* a valid solution:

> Let n be the number of balls thrown. This problem is equivalent to the birthday paradox, where b is the number of days in the year, n is the number of people in the room, and two balls share a bin if and only if two people share a birthday. So, if X is the number of pairs of balls that share a bin, then (following the discussion in CLRS of the birthday paradox) we have that
>
> $$\mathrm{E}[X] = \frac{n(n-1)}{2b}$$
>
> and $\mathrm{E}[X] \geq 1$ when $n \geq \sqrt{2b} + 1$.

To see why, note that the ball tosses should stop as soon as two balls collide. Here, $\mathrm{E}[X]$ includes outcomes where more than two balls land in a bin.

(5.4-6) $\star$

Suppose that n balls are tossed into n bins, where each toss is independent and the ball is equally likely to end up in any bin. What is the expected number of empty bins? What is the expected number of bins with exactly one ball?

Let X be the number of empty bins. We'll use indicator variables to find $\mathrm{E}[X]$:

$$X_i := \text{"after } n \text{ throws, the } i\text{th bin is empty"} \quad .$$

Since X_i is an indicator variable, $\mathrm{E}[X_i] = \Pr[X_i]$, and by linearity of expectation, $E[X] = \sum_{i=1}^{n} \Pr[X_i]$. What is $\Pr[X_i]$? If we throw a single ball, let p be the probability that it lands in bin i, and let $q = 1 - p$ be the probability that it doesn't. Bin i is empty if and only if n such throws failed to land in bin i. These are n independent Bernoulli trials with p and q, so we have that

$$\Pr[X_i] = \binom{n}{n} p^0 q^n = (1-p)^n = \left(1 - \frac{1}{n}\right)^n$$

and

$$\mathrm{E}[X] = \sum_{i=1}^{n} \Pr[X_i] = n\left(1 - \frac{1}{n}\right)^n \quad .$$

Recalling that $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$, we have $\mathrm{E}[X] \approx \frac{n}{e}$ for large n.

Now let X be the number of bins with exactly one ball, and let

$$X_i := \text{"after } n \text{ throws, the } i\text{th bin has exactly one ball"} \quad .$$

We have that $E[X] = \sum_{i=1}^{n} \Pr[X_i]$, so let's find $\Pr[X_i]$. As before, let p be the probability that a throw lands in bin i, and let $q = 1 - p$. The probability that bin i has exactly one ball is the probability that exactly $n-1$ balls failed to land in bin i, or exactly $n-1$ failed trials. We have that

$$\begin{aligned}
\Pr[X_i] &= \binom{n}{1} p^1 q^{n-1} = n\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} = \left(1-\frac{1}{n}\right)^{n-1} \\
\mathrm{E}[X] &= \sum_{i=1}^{n} \Pr[X_i] = n\left(1-\frac{1}{n}\right)^{n-1} \\
&= n\frac{\left(1-\frac{1}{n}\right)^{n-1}\left(1-\frac{1}{n}\right)}{\left(1-\frac{1}{n}\right)} = n\frac{\left(1-\frac{1}{n}\right)^{n}}{\left(1-\frac{1}{n}\right)} = n^2\frac{\left(1-\frac{1}{n}\right)^{n}}{(n-1)}\ , \\
\lim_{n\to\infty} \mathrm{E}[X] &= \frac{n^2}{e(n-1)}\ .
\end{aligned}$$

(5-1) Probabilistic counting

With a b-bit counter, we can ordinarily only count up to $2^b - 1$. With R. Morris's ***probabilistic counting****, we can count up to a much larger value at the expense of some loss of precision.*

We let a counter value of i represent a count of n_i for $i = 0, 1, \ldots, 2^b - 1$, where the n_i form an increasing sequence of nonnegative values. We assume that the initial value of the counter is 0, representing a count of $n_0 = 0$. The INCREMENT *operation works on a counter containing the value i in a probabilistic manner. If $i = 2^b - 1$, then an overflow error is reported. Otherwise, the counter is increased by 1 with probability $1/(n_{i+1} - n_i)$, and it remains unchanged with probability $1 - 1/(n_{i+1} - n_i)$.*

If we select $n_i = i$ for all $i \geq 0$, then the counter is an ordinary one. More interesting situations arise if we select, say, $n_i = 2^{i-1}$ for $i > 0$ or $n_i = F_i$ (the ith Fibonacci number – see Section 3.2).

For this problem, assume that n_{2^b-1} is large enough that the probability of an overflow error is negligible.

(a). *Show that the expected value represented by the counter after n* INCREMENT *operations have been performed is exactly n.*

Let N_k be the value represented by the counter after k operations. N_k will be n_i for some i. After the next operation, it has expected value

$$\begin{aligned}
\mathrm{E}[N_{k+1} \mid N_k] &= \frac{1}{n_{i+1}-n_i} n_{i+1} + \left(1 - \frac{1}{n_{i+1}-n_i}\right) n_i \\
&= \frac{n_{i+1}-n_i}{n_{i+1}-n_i} + n_i \\
&= 1 + n_i = 1 + N_k\ .
\end{aligned}$$

We can then use induction. For the base case, $\mathrm{E}[N_0] = n_0 = 0$. Assume $\mathrm{E}[N_k] = k$. Then by iterated expectation

$$\mathrm{E}[N_{k+1}] = \mathrm{E}[\mathrm{E}[N_{k+1} \mid N_k]] = \mathrm{E}[1 + N_k] = 1 + \mathrm{E}[N_k] = 1 + k\ .$$

(b). *The analysis of the variance of the count represented by the counter depends on the sequence of the n_i. Let us consider a simple case: $n_i = 100i$ for all $i \geq 0$. Estimate the variance in the value represented by the register after n* Increment *operations have been performed.*

Consider the probability p that we increase a counter value from n_i to n_{i+1}:

$$p = \frac{1}{n_{i+1} - n_i} = \frac{1}{100i + 1 - 100i} = \frac{1}{100}$$

Since this value is constant for all i, we can view counter updates as Bernoulli trials. Let I be the number of times that the counter is increased in n Increment operations. Since I is binomially distributed, it has variance

$$\mathrm{Var}[I] = np(1-p) = \frac{99n}{100^2} ,$$

and the variance of the value n_I represented by the counter is

$$\mathrm{Var}[n_I] = \mathrm{Var}[100I] = 100^2 \mathrm{Var}[I] = 99n .$$

Iterated Monte Carlo algorithm

Say we have a Monte-carlo randomized algorithm A which may return the wrong answer, but the $\Pr[A \textit{ is wrong}] \le \frac{1}{n^\beta}$*, where n is the size of the input and $\beta > 0$ is a constant. How many repetitions of algorithm A are required to reduce the error probability to $1/2^n$?*

We are given that $\Pr[A \text{ is wrong}] \le \frac{1}{n^\beta}$. Assuming independence, then after k repetitions

$$\Pr[A \text{ is wrong after } k \text{ repetitions}] \le \left(\frac{1}{n^\beta}\right)^k .$$

We'd like to find the minimum k such that $\left(\frac{1}{n^\beta}\right)^k \le \frac{1}{2^n}$. Rewriting as $n^{k\beta} \ge 2^n$ and solving for k, we find that the error probability is reduced to $1/2^n$ if

$$k \ge \frac{n}{\beta \lg n} .$$