Exercise 1. Let $A(m, n)$ denote the number of configurations of putting m balls numbered from 1 to m into n bins such that no bin is empty. So, the probability that there is no empty bin after throwing m balls into n bins is $A(m, n)/n^m$.

1. Show that $A(m, n+1) = \sum_{r=1}^{m} \binom{m}{r} A(m-r, n)$.
2. Show by induction that $A(m, n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^m$.
3. Let $E_k(m, n)$ be the number of arrangements leaving exactly k bins empty. Show that $E_k(m, n) = \binom{n}{k} A(m, n-k)$, and find a closed form formula for this quantity.

solution :

1. We show this formula by partitioning the space according to the number r of balls into the last bin. If $r = 0$ a bin is empty so no such configuration counts in $A(m, n+1)$. For a given r, there are $\binom{m}{r}$ ways to choose the r balls that will end up in the last bin, and $A(m-r, n)$ ways for the other balls to leave no empty bin among the first n bin.
2. Let us prove the formula by induction on (n, m). For all m we have $A(m, 1) = 1$ and for all n we have $A(0, n) = 0$. We suppose the result correct for all $(a, b) < (m, n+1)$ and show it is true for $A(m, n+1)$. Using (1) we have

$$A(m, n+1) = \sum_{r=1}^{m} \binom{m}{r} A(m-r, n)$$

and by induction hypothesis

$$A(m, n+1) = \sum_{r=1}^{m} \binom{m}{r} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^{m-r} .$$

We permute the sum,

$$A(m, n+1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left[\sum_{r=1}^{m} \binom{m}{r} (n-k)^{m-r} \right] .$$

The term in brackets is almost $(n+1-k)^m$, so we make it appear :

$$A(m, n+1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left[(n+1-k)^m - (n-k)^m \right] .$$

We now split the sum in two to regroup the same term under the power of m :

$$A(m, n+1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n+1-k)^m - \sum_{y=1}^{n+1} (-1)^{y-1} \binom{n}{(y-1)} (n-(y-1))^m .$$

We obtain

$$A(m, n+1) = (-1)^0 \binom{n}{0} (n+1-0)^m + \sum_{k=1}^{n} (n+1-k)^m \left[(-1)^k \binom{n}{k} - (-1)^{k-1} \binom{n}{k-1} \right] - (-1)^n \binom{n}{n} (0)^m ,$$

that is

$$A(m, n+1) = (-1)^0 \binom{n+1}{0} (n+1-0)^m + \sum_{k=1}^{n} (-1)^k (n+1-k)^m \left[\binom{n}{k} + \binom{n}{k-1} \right] + (-1)^{n+1} \binom{n+1}{n+1} (0)^m .$$

The term in bracket is nothing but $\binom{n+1}{k}$ so we get the result.

3. We have $\binom{n}{k}$ ways to choose the empty bins, and then exactly $A(m, n-k)$ to fill the others without leaving any empty bin, hence the result. Using (2) this quantity is equal to

$$E_k(m,n) = \binom{n}{k} \sum_{i=0}^{n-k} (-1)^i \binom{n-k}{i} (n-k-i)^m$$

Exercise 2. Let D be a probability distribution on the numbers $\{1, ..., m\}$, and assume that the probability of i is p_i. Let $p(x) = \sum_{i=1}^{m} p_i x^i$ be the generating function of the distribution.

1. Let X be a random variable with distribution D. Show that $E[X] = p'(1)$, where $p'(x)$ is the derivative of $p(x)$.
2. Similarly, find a formula for $\text{Var}[X]$ as a linear combination of the first and the second derivatives of $p(x)$.
3. Let X be geometrically distributed with mean p, i.e., $Pr[X = t] = p(1-p)^{t-1}$. Find the generating function of this distribution, its mean, and its variance.

solution (not complete):

1. we have

$$p'(x) = \sum_{i=1}^{m} i p_i x^{i-1}$$

$$E[X] = \sum_{i=1}^{m} i p_i = p'(1)$$

2. we have

$$\text{Var}[X] = E[X^2] - E[X]^2 = \sum i^2 p_i - p'(1)^2 = [xp'(x)]'(1) - p'(1)^2$$

$$\text{Var}[X] = [p'(x) + xp''(x)](1) - p'(1)^2 = p'(1) + p''(1) - p'(1)^2 \ .$$

Exercise 3 (Power of two choices)**.** We want to prove here that in the balls and bins problem, if we pick two bins at random and place the ball in the bin with fewer balls, then the maximum load will be of the order of $O(ln(ln(n)))$, when throwing n balls into n bins.

We write $N_i(n)$ for the number of bins with at least i balls after throwing n balls. The idea of the proof is to introduce some number β_i such that $N_i(n) \le \beta_i$ holds with high probability.

1. Show that if $\beta_2 = \frac{n}{2}$, then $\Pr(N_2(n) \le \beta_2)$ is 1.
2. For a given number a, we want to compute

$$\Pr(N_{i+1}(n) > a \mid N_i(n) \le \beta_i) \ .$$

For this, assuming that at any point of time there is no more than β_i bins with more than or exactly i balls, what is an upper bound on the probability that a given ball ends up in such a bin ? Using a Bernoulli law and a Chernoff type bound, show that with

$$\beta_{i+1}/n = (1+\delta)(\beta_i/n)^2 \ ,$$

$$\Pr(N_{i+1}(n) > \beta_{i+1} \mid N_i(n) \le \beta_i) \le e^{-\frac{\delta^2 \beta_i^2}{2n}} \ .$$

3. If we choose $\delta = \sqrt{2}$ this probability is smaller than $\frac{1}{n^2}$ as long as $\beta_i > \sqrt{2n\log(n)}$. Let i^* be the first i such that this is not the case. What is this i^* ? we set $\beta_{i^*} = \sqrt{2n\log(n)}$. With

$$\beta_{i^*+1}/n = (1+\sqrt{2})(\beta_{i^*}/n)^2 = (1+\sqrt{2})2\log n/n$$

we still have

$$\Pr(N_{i^*+1}(n) > \beta_{i^*+1} \mid N_{i^*}(n) \le \beta_{i^*}) \le \frac{1}{n^2} \ .$$

Finally, what δ should we choose to have $\beta_{i^*+2} < 1$ and

$$\Pr(N_{i^*+2}(n) \ge 1 \mid N_{i^*+1}(n) \le \beta_{i^*+1}) \le \frac{1}{n^2} \ ?$$

4. To conclude, using that $P(\neg A) \le P(\neg A|B) + P(\neg B)$ show that if $i^* = \log(\log(n))/\log(2) + O(1)$ then

$$\Pr(N_{i^*+2}(n) \ge 1) \le \frac{i^*}{n^2} \ .$$