Using Stirling numbers to solve coupon collector problems

Marko R. Riedel

October 28, 2025

The coupon collector problem has been studied in many variations, from basic probability to advanced research. For an introduction consult the Wikipedia entry [Wik17] listed in the references. Some results here use the Egorychev method from [Ego84] to compute binomial coefficient sums by residues.

We present five calculations ranging in difficulty from beginner to more advanced. The Egorychev method (summation by residues of rational functions) is used to evaluate some of the more difficult sums that appear. This document has retained the question answer format used by Math Stackexchange.

Contents

1 Basic example, solving the standard problem by Stirling numbers

We start with a proof using Stirling numbers of the second kind which encapsulates inclusion-exclusion in the generating function of these numbers. First let us verify that we indeed have a probability distribution here. We have for the number T of coupons being m draws that

$$P[T = m] = \frac{1}{n^m} \times n \times \left\{ {m-1 \atop n-1} \right\} \times (n-1)!.$$

Recall the OGF of the Stirling numbers of the second kind which says that

$$\left\{ {n \atop k} \right\} = [z^n] \prod_{q=1}^{k} \frac{z}{1-qz}.$$

This gives for the sum of the probabilities

$$\sum_{m \ge 1} P[T = m] = (n-1)! \sum_{m \ge 1} \frac{1}{n^{m-1}} \left\{ {m-1 \atop n-1} \right\}$$

$$= (n-1)! \sum_{m \ge 1} \frac{1}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{n-1} \frac{z}{1-qz}$$

$$= (n-1)! \prod_{q=1}^{n-1} \frac{1/n}{1-q/n} = (n-1)! \prod_{q=1}^{n-1} \frac{1}{n-q} = 1.$$

This confirms it being a probability distribution.
We then get for the expectation that

$$\sum_{m \ge 1} m \times P[T = m] = (n-1)! \sum_{m \ge 1} \frac{m}{n^{m-1}} \left\{ {m-1 \atop n-1} \right\}$$

$$= (n-1)! \sum_{m \ge 1} \frac{m}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{n-1} \frac{z}{1-qz}$$

$$= 1 + (n-1)! \sum_{m \ge 1} \frac{m-1}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{n-1} \frac{z}{1-qz}$$

$$= 1 + (n-1)! \sum_{m \ge 2} \frac{m-1}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{n-1} \frac{z}{1-qz}$$

$$= 1 + \frac{1}{n}(n-1)! \sum_{m \ge 2} \frac{1}{n^{m-2}} [z^{m-2}] \left(\prod_{q=1}^{n-1} \frac{z}{1-qz} \right)'$$

$$= 1 + \frac{1}{n}(n-1)! \left(\prod_{q=1}^{n-1} \frac{z}{1-qz} \right)' \Bigg|_{z=1/n}$$

$$= 1 + \frac{1}{n}(n-1)! \left(\prod_{q=1}^{n-1} \frac{z}{1-qz} \sum_{p=1}^{n-1} \frac{1-pz}{z} \frac{1}{(1-pz)^2} \right) \Bigg|_{z=1/n}$$

$$= 1 + \frac{1}{n}(n-1)! \prod_{q=1}^{n-1} \frac{1/n}{1-q/n} \sum_{p=1}^{n-1} \frac{1}{z} \frac{1}{1-pz} \Bigg|_{z=1/n}$$

$$= 1 + \frac{1}{n}(n-1)! \prod_{q=1}^{n-1} \frac{1}{n-q} \sum_{p=1}^{n-1} \frac{n}{1-p/n}$$

$$= 1 + \frac{1}{n} \sum_{p=1}^{n-1} \frac{n^2}{n-p} = 1 + nH_{n-1} = n \times H_n.$$

This was math.stackexchange.com problem 1609459.

2 Number of tuples of some size seen

To start let us state the probability of needing m draws to collect all n coupons. Using combinatorial classes as in *Analytic Combinatorics* we get with the labeled set operator / class

$$\mathrm{P}[T = m] = \frac{1}{n^m} \times \binom{n}{1} \times (m-1)![z^{m-1}](\exp(z)-1)^{n-1}.$$

Let us just verify that this will produce $n \times H_n$ for the number of draws. We get

$$\mathrm{E}[T] = \sum_{m \geq 1} \frac{nm}{n^m}(m-1)![z^{m-1}](\exp(z)-1)^{n-1}$$

$$= \sum_{m \geq 0} \frac{m+1}{n^m} m![z^m](\exp(z)-1)^{n-1}$$

$$= \sum_{m \geq 0} \frac{m+1}{n^m} m![z^m] \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \exp(qz)$$

$$= \sum_{m \geq 0} \frac{m+1}{n^m} \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} q^m$$

$$= \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \sum_{m \geq 0} \frac{m+1}{n^m} q^m$$

$$= \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \frac{1}{(1-q/n)^2}.$$

Now to evaluate this last sum we introduce

$$f(z) = (n-1)! \frac{1}{(1-z/n)^2} \prod_{r=0}^{n-1} \frac{1}{z-r} = n! \frac{n}{(z-n)^2} \prod_{r=0}^{n-1} \frac{1}{z-r}.$$

which has the property that

$$\operatorname*{res}_{z=q} f(z) = \binom{n-1}{q} (-1)^{n-1-q} \frac{1}{(1-q/n)^2}.$$

Residues sum to zero and we may evaluate with minus the residue at $z = n$:

$$-n! \times n \times \left. \left(\prod_{r=0}^{n-1} \frac{1}{z-r} \right)' \right|_{z=n}$$

$$= n! \times n \times \left. \left(\prod_{r=0}^{n-1} \frac{1}{z-r} \sum_{r=0}^{n-1} \frac{1}{z-r} \right) \right|_{z=n} = n! \times n \times \frac{1}{n!} H_n = n \times H_n.$$

The sanity check goes through.

Main computation

Next we ask how many coupons appear j times in this setting. We get the mixed GF

$$\frac{1}{n^m} \times \binom{n}{1} \times (m-1)! [z^{m-1}] \left(\exp(z) + (u-1) \frac{z^j}{j!} - 1 \right)^{n-1}.$$

Differentiate with respect to u and set $u = 1$ to get

$$\frac{1}{n^m} \times n(n-1) \times (m-1)! [z^{m-1}] (\exp(z) - 1)^{n-2} \frac{z^j}{j!}.$$

Summing we find

$$\frac{n(n-1)}{j!} \sum_{m \ge j+1} \frac{1}{n^m} (m-1)! [z^{m-1-j}] (\exp(z) - 1)^{n-2}$$

$$= \frac{n(n-1)}{j!} \sum_{m \ge j+1} \frac{1}{n^m} (m-1)! [z^{m-1-j}] \sum_{q=0}^{n-2} \binom{n-2}{q} (-1)^{n-q} \exp(qz)$$

$$= n(n-1) \sum_{m \ge j+1} \frac{1}{n^m} \binom{m-1}{m-1-j} \sum_{q=0}^{n-2} \binom{n-2}{q} (-1)^{n-q} q^{m-1-j}$$

$$= (n-1)\frac{1}{n^j}\sum_{m\ge 0}\frac{1}{n^m}\binom{m+j}{m}\sum_{q=0}^{n-2}\binom{n-2}{q}(-1)^{n-q}q^m$$

$$= (n-1)\frac{1}{n^j}\sum_{q=0}^{n-2}\binom{n-2}{q}(-1)^{n-q}\frac{1}{(1-q/n)^{j+1}}$$

$$= (-1)^{j+1}n(n-1)\sum_{q=0}^{n-2}\binom{n-2}{q}(-1)^{n-q}\frac{1}{(q-n)^{j+1}}.$$

To evaluate we introduce

$$f(z) = (n-2)!\frac{1}{(z-n)^{j+1}}\prod_{r=0}^{n-2}\frac{1}{z-r}$$

which has the property that

$$\operatorname*{res}_{z=q} f(z) = \binom{n-2}{q}(-1)^{n-q}\frac{1}{(q-n)^{j+1}}.$$

Residues sum to zero and we may evaluate with minus the residue at $z = n$ since the residue at infinity is zero also, and we obtain in terms of a derivative

$$\boxed{\mathrm{E}[C_j] = (-1)^j\frac{n!}{j!}\left[\prod_{r=0}^{n-2}\frac{1}{z-r}\right]^{(j)}\Bigg|_{z=n.}}$$

We can use this formula to get concrete values for fixed j. We find for singletons with $j = 1$

$$-n!\left[\prod_{r=0}^{n-2}\frac{1}{z-r}\sum_{r=0}^{n-2}\frac{1}{r-z}\right]\Bigg|_{z=n.} = -n!\times\frac{1}{n!}\times(-H_n+1)$$

so that

$$\boxed{\mathrm{E}[C_1] = H_n.}$$

Here we have added a unique correction term as the coupon that completes the set also counts as a singleton.

Continuing with pairs, i.e. $j = 2$,

$$\frac{1}{2}n!\left[\prod_{r=0}^{n-2}\frac{1}{z-r}\left(\sum_{r=0}^{n-2}\frac{1}{r-z}\right)^2 + \prod_{r=0}^{n-2}\frac{1}{z-r}\sum_{r=0}^{n-2}\frac{1}{(r-z)^2}\right]\Bigg|_{z=n.}$$

so that we obtain $\frac{1}{2}[(H_n-1)^2 + H_n^{(2)} - 1]$ or

$$\boxed{\mathrm{E}[C_2] = \frac{1}{2}[H_n^2 - 2H_n + H_n^{(2)}].}$$

To conclude we do triples i.e. $j = 3$. We need to differentiate

$$\prod_{r=0}^{n-2} \frac{1}{z-r} \left[\left(\sum_{r=0}^{n-2} \frac{1}{r-z} \right)^2 + \sum_{r=0}^{n-2} \frac{1}{(r-z)^2} \right]$$

which gives

$$\prod_{r=0}^{n-2} \frac{1}{z-r} \sum_{r=0}^{n-2} \frac{1}{r-z} \left[\left(\sum_{r=0}^{n-2} \frac{1}{r-z} \right)^2 + \sum_{r=0}^{n-2} \frac{1}{(r-z)^2} \right]$$
$$+ \prod_{r=0}^{n-2} \frac{1}{z-r} \left[\left(2 \sum_{r=0}^{n-2} \frac{1}{r-z} \sum_{r=0}^{n-2} \frac{1}{(r-z)^2} \right) + 2 \sum_{r=0}^{n-2} \frac{1}{(r-z)^3} \right]$$

Setting $z = n$ we get

$$-\frac{1}{6}(-(H_n - 1)[(H_n - 1)^2 + H_n^{(2)} - 1] - 2(H_n - 1)(H_n^{(2)} - 1) - 2(H_n^{(3)} - 1))$$

or

$$\boxed{\mathrm{E}[C_3] = \frac{1}{6}((H_n - 1)[(H_n - 1)^2 + 3(H_n^{(2)} - 1)] + 2(H_n^{(3)} - 1)).}$$

Asymptotics

For the asymptotics we get

$$\boxed{\mathrm{E}[C_1] \sim \log n + \gamma.}$$

as well as for $j = 2$

$$\frac{1}{2}[\log^2 n + 2\gamma \log n + \gamma^2 - 2 \log n - 2\gamma + \frac{\pi^2}{6}]$$

which is

$$\boxed{\mathrm{E}[C_2] \sim \frac{1}{2} \log^2 n + (\gamma - 1) \log n + \frac{1}{2}\gamma^2 - \gamma + \frac{\pi^2}{12}.}$$

The case for $j = 3$ is left to the reader. Apery's constant appears.

Computer Algebra

Using a CAS to automate the differentiation in the formula for $\mathrm{E}[C_j]$ we can tackle higher values of j e.g. $j = 20$ is possible. We get for $j = 4$

$$\begin{aligned}\mathrm{E}[C_4] = {} & \frac{1}{24} H_{n,1}^4 - \frac{1}{6} H_{n,1}^3 + \frac{1}{4} H_{n,1}^2 H_{n,2} - \frac{1}{2} H_{n,1} H_{n,2} \\ & + \frac{1}{3} H_{n,1} H_{n,3} - \frac{1}{3} H_{n,3} + \frac{1}{8} H_{n,2}^2 + \frac{1}{4} H_{n,4}\end{aligned}$$

and for $j = 5$

$$\begin{aligned}\mathrm{E}[C_5] = {} & \frac{1}{6} H_{n,2} H_{n,3} + \frac{1}{12} H_{n,1}^3 H_{n,2} + \frac{1}{6} H_{n,1}^2 H_{n,3} + \frac{1}{8} H_{n,1} H_{n,2}^2 + \frac{1}{4} H_{n,1} H_{n,4} \\ & + \frac{1}{5} H_{n,5} + \frac{1}{120} H_{n,1}^5 - \frac{1}{3} H_{n,1} H_{n,3} - \frac{1}{8} H_{n,2}^2 \\ & - \frac{1}{4} H_{n,1}^2 H_{n,2} - \frac{1}{4} H_{n,4} - \frac{1}{24} H_{n,1}^4.\end{aligned}$$

We have for the asymptotics for $j = 3$

$$\begin{aligned}\mathrm{E}[C_3] \sim {} & \frac{\ln(n)^3}{6} + \left(\frac{\gamma}{2} - \frac{1}{2}\right) \ln(n)^2 + \left(\frac{1}{2}\gamma^2 - \gamma + \frac{1}{12}\pi^2\right) \ln(n) \\ & + \frac{\gamma^3}{6} - \frac{\gamma^2}{2} + \frac{\gamma\,\pi^2}{12} - \frac{\pi^2}{12} + \frac{\zeta(3)}{3}.\end{aligned}$$

and for $j = 4$

$$\begin{aligned}\mathrm{E}[C_4] \sim {} & \frac{\ln(n)^4}{24} + \left(-\frac{1}{6} + \frac{\gamma}{6}\right) \ln(n)^3 + \left(-\frac{1}{2}\gamma + \frac{1}{4}\gamma^2 + \frac{1}{24}\pi^2\right) \ln(n)^2 \\ & + \left(\frac{\gamma^3}{6} - \frac{\gamma^2}{2} + \frac{\gamma\,\pi^2}{12} - \frac{\pi^2}{12} + \frac{\zeta(3)}{3}\right) \ln(n) \\ & + \frac{\pi^4}{160} - \frac{\zeta(3)}{3} - \frac{\gamma\,\pi^2}{12} - \frac{\gamma^3}{6} + \frac{\gamma^2\pi^2}{24} + \frac{\gamma\zeta(3)}{3} + \frac{\gamma^4}{24}.\end{aligned}$$

Observe that adding $\sum_{j\geq 1} j \times \frac{1}{j!} \log^j(n) \sim \log(n) \times \exp(\log n) = n \times \log(n)$ which shows we have accounted for all coupons.

This was math.stackexchange.com problem 4970981.

3 Drawing coupons until at least j instances of each type are seen

What follows is a minor contribution where we compute a formula for the expectation for the case where j instances of each of n types of coupons must be seen. Using the notation from the following MSE link we have from first principles that

$$P[T = m] = \frac{1}{n^m} \times \binom{n}{1} \times (m-1)![z^{m-1}] \left(\exp(z) - \sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1} \frac{z^{j-1}}{(j-1)!}$$

$$= \frac{n}{n^m} \times \binom{m-1}{j-1} (m-j)![z^{m-j}] \left(\exp(z) - \sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1}.$$

We **verify that this is a probability distribution**. The goal here is to find a closed form for the infinite series in m so that its value may be calculated rather than approximated. Expanding the power we find

$$\sum_{m \ge j} P[T = m] = \frac{n}{n^j} \sum_{m \ge 0} \frac{1}{n^m} \times \binom{m+j-1}{j-1} m![z^m] \sum_{k=0}^{n-1} \binom{n-1}{k} \exp(kz)$$

$$\times (-1)^{n-1-k} \left(\sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1-k}$$

$$= \frac{n}{n^j} \sum_{m \ge 0} \frac{1}{n^m} \times \binom{m+j-1}{j-1} m! \sum_{k=0}^{n-1} \binom{n-1}{k} \sum_{p=0}^{m} \frac{k^{m-p}}{(m-p)!}$$

$$\times (-1)^{n-1-k} [z^p] \left(\sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1-k}$$

$$= \frac{n}{n^j} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^{n-1-k} \sum_{p \ge 0} [z^p] \left(\sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1-k}$$

$$\times \sum_{m \ge p} \frac{1}{n^m} \times \binom{m+j-1}{j-1} m! \times \frac{k^{m-p}}{(m-p)!}$$

$$= \frac{n}{n^j} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^{n-1-k} \sum_{p \ge 0} n^{-p} [z^p] \left(\sum_{q=0}^{j-1} \frac{z^q}{q!}\right)^{n-1-k}$$

$$\times \sum_{m \ge 0} \frac{1}{n^m} \times \binom{m+p+j-1}{j-1} (m+p)! \times \frac{k^m}{m!}$$

The inner sum is

$$\frac{1}{(j-1)!}\sum_{m\ge 0}(k/n)^m\frac{(m+p+j-1)!}{m!} = \frac{(p+j-1)!}{(j-1)!}\frac{1}{(1-k/n)^{p+j}}$$

and with $P=(n-1-k)(j-1)$ we obtain

$$n\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{(-1)^{n-1-k}}{(n-k)^j}\times\sum_{p=0}^{P}\frac{1}{(n-k)^p}\frac{(p+j-1)!}{(j-1)!}[z^p]\left(\sum_{q=0}^{j-1}\frac{z^q}{q!}\right)^{n-1-k}.$$

We claim that the inner sum is $(n-k)^{j-1}$, proof for $j=2$ at end of document. With this the sum reduces to

$$n\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{(-1)^{n-1-k}}{n-k}$$
$$= -\sum_{k=0}^{n-1}\binom{n}{k}(-1)^{n-k} = 1-\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k} = 1$$

and we see that we indeed have a probability distribution.

Continuing with the expectation and re-capitulating the earlier computation we find

$$E[T] = \sum_{m\ge j} mP[T=m] = \frac{n}{n^j}\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{n-1-k}\sum_{p\ge 0}n^{-p}[z^p]\left(\sum_{q=0}^{j-1}\frac{z^q}{q!}\right)^{n-1-k}$$
$$\times\sum_{m\ge 0}\frac{1}{n^m}\times\binom{m+p+j-1}{j-1}(m+p)!(m+p+j)\times\frac{k^m}{m!}$$

The inner sum has two pieces, the first is

$$\frac{1}{(j-1)!}\sum_{m\ge 1}(k/n)^m\frac{(m+p+j-1)!}{(m-1)!} = \frac{1}{(j-1)!}\frac{k}{n}\sum_{m\ge 0}(k/n)^m\frac{(m+p+j)!}{m!}$$
$$= \frac{k}{n}\frac{(p+j)!}{(j-1)!}\frac{1}{(1-k/n)^{p+j+1}} = \frac{k}{n-k}\frac{(p+j)!}{(j-1)!}\frac{1}{(1-k/n)^{p+j}}$$

and the second has been evaluated when we summed the probabilities to give

$$(p+j)\frac{(p+j-1)!}{(j-1)!}\frac{1}{(1-k/n)^{p+j}} = \frac{(p+j)!}{(j-1)!}\frac{1}{(1-k/n)^{p+j}}.$$

Substituting these into the outer sum we thus obtain

$$E[T] = n^2 \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{(-1)^{n-1-k}}{(n-k)^{j+1}} \times \sum_{p=0}^{P} \frac{1}{(n-k)^p} \frac{(p+j)!}{(j-1)!} [z^p] \left(\sum_{q=0}^{j-1} \frac{z^q}{q!} \right)^{n-1-k}.$$

There is a very basic program which confirmed this formula for several digits of precision by simulation which is written in C and goes as follows.

Proof of inner sum for $j = 2$. Setting $j = 2$ we have to show that

$$n - k = \sum_{p=0}^{n-1-k} \binom{n-1-k}{p} \frac{1}{(n-k)^p} (p+1)!.$$

This is

$$\frac{1}{(n-1-k)!} = \sum_{p=0}^{n-1-k} \frac{1}{(n-1-k-p)!} \frac{p+1}{(n-k)^{p+1}}.$$

Re-writing as follows

$$\frac{1}{m!} = \sum_{p=0}^{m} \frac{1}{(m-p)!} \frac{p+1}{(m+1)^{p+1}}.$$

and introducing

$$\frac{1}{(m-p)!} = \frac{1}{2\pi i} \int_{|w|=\gamma} \frac{1}{w^{m-p+1}} \exp(w)\, dw$$

we obtain for the sum (the integral vanishes nicely when $p > m$ so we may extend p to infinity)

$$\frac{1}{2\pi i} \int_{|w|=\gamma} \frac{1}{w^{m+1}} \exp(w) \frac{1}{m+1} \sum_{p \geq 0} \frac{p+1}{(m+1)^p} w^p \, dw$$
$$= \frac{1}{m+1} \frac{1}{2\pi i} \int_{|w|=\gamma} \frac{1}{w^{m+1}} \exp(w) \frac{1}{(1 - w/(m+1))^2} \, dw.$$

We now use the fact that residues sum to zero and the poles are at $w = 0$, $w = m+1$ and $w = \infty$. We get for the residue at infinity

$$-\frac{1}{m+1} \mathrm{Res}_{w=0} \frac{1}{w^2} w^{m+1} \exp(1/w) \frac{1}{(1 - 1/w/(m+1))^2}$$
$$= -\frac{1}{m+1} \mathrm{Res}_{w=0} w^{m+1} \exp(1/w) \frac{1}{(w - 1/(m+1))^2}$$
$$= -(m+1) \mathrm{Res}_{w=0} w^{m+1} \exp(1/w) \frac{1}{(1 - w(m+1))^2}$$

$$= -(m+1)[w^{-(m+2)}]\exp(1/w)\frac{1}{(1-w(m+1))^2}.$$

Extracting coefficients we find

$$-(m+1)\sum_{q\ge 0}\frac{1}{(q+m+2)!}(q+1)(m+1)^q.$$

This is

$$-(m+1)\left(\sum_{q\ge 0}\frac{1}{(q+m+1)!}(m+1)^q - \sum_{q\ge 0}\frac{1}{(q+m+2)!}(m+1)^{q+1}\right)$$

$$= -(m+1)\left(\sum_{q\ge 0}\frac{1}{(q+m+1)!}(m+1)^q - \sum_{q\ge 1}\frac{1}{(q+m+1)!}(m+1)^q\right)$$

$$= -(m+1)\frac{(m+1)^0}{(m+1)!} = -\frac{1}{m!}.$$

We thus have the claim if we can show the residue at $w = m+1$ is zero. We use

$$(m+1)\frac{1}{2\pi i}\int_{|w|=\gamma}\frac{1}{w^{m+1}}\exp(w)\frac{1}{(w-(m+1))^2}\,dw.$$

and observe that

$$\left.\left(\frac{1}{w^{m+1}}\exp(w)\right)'\right|_{w=m+1}$$

$$= \left.\left(-(m+1)\frac{1}{w^{m+2}}\exp(w) + \frac{1}{w^{m+1}}\exp(w)\right)\right|_{w=m+1}$$

$$= \exp(m+1)\left(-(m+1)\frac{1}{(m+1)^{m+2}} + \frac{1}{(m+1)^{m+1}}\right) = 0$$

as required. This concludes the computation.
This was math.stackexchange.com problem 2426510.

3.1 Drawing coupons until at least 2 instances of each type are seen with n' types already collected

What follows is a computational contribution where we derive a closed form (as opposed to an infinite series) of the expected number of draws required to see all coupons at least twice when a number n' of coupons from the n types where $n' < n$ have already been collected in two instances. We then observe that the expectation does not simplify. It seems like a rewarding challenge to compute the asymptotics for these expectations using probabilistic methods and compare them to the closed form presented below.

Using the notation from this MSE link we have from first principles that

$$P[T=m] = \frac{1}{n^m} \times {n-n' \choose 1} \times (m-1)! [z^{m-1}] \exp(n'z) \left(\exp(z)-1-z\right)^{n-n'-1} \frac{z}{1}.$$

We **verify that this is a probability distribution**. We get

$$\sum_{m\ge 2} P[T=m]$$

$$= (n-n') \sum_{m\ge 2} \frac{1}{n^m} (m-1)! [z^{m-2}] \exp(n'z) \left(\exp(z)-1-z\right)^{n-n'-1}$$

$$= (n-n') \frac{1}{n^2} \sum_{m\ge 0} \frac{1}{n^m} (m+1)! [z^m] \exp(n'z) \left(\exp(z)-1-z\right)^{n-n'-1}$$

$$= (n-n') \frac{1}{n^2} \sum_{m\ge 0} \frac{1}{n^m} (m+1)! [z^m] \exp(n'z)$$

$$\times \sum_{p=0}^{n-n'-1} {n-n'-1 \choose p} \exp((n-n'-1-p)z) (-1)^p (1+z)^p$$

$$= (n-n') \frac{1}{n^2} \sum_{m\ge 0} \frac{1}{n^m} (m+1)!$$

$$\times [z^m] \sum_{p=0}^{n-n'-1} {n-n'-1 \choose p} \exp((n-1-p)z) (-1)^p (1+z)^p$$

$$= (n-n') \frac{1}{n^2} \sum_{m\ge 0} \frac{1}{n^m} (m+1)!$$

$$\times \sum_{p=0}^{n-n'-1} {n-n'-1 \choose p} \sum_{q=0}^{m} [z^{m-q}] \exp((n-1-p)z) (-1)^p [z^q] (1+z)^p$$

$$= (n-n') \frac{1}{n^2} \sum_{m\ge 0} \frac{1}{n^m} (m+1)!$$

$$\times \sum_{p=0}^{n-n'-1} {n-n'-1 \choose p} \sum_{q=0}^{m} \frac{(n-1-p)^{m-q}}{(m-q)!} (-1)^p {p \choose q}.$$

Re-arranging the order of the sums now yields

$$(n-n') \frac{1}{n^2} \sum_{p=0}^{n-n'-1} {n-n'-1 \choose p}$$

$$\times \sum_{m\ge 0} \frac{1}{n^m}(m+1)! \sum_{q=0}^{m} \frac{(n-1-p)^{m-q}}{(m-q)!} (-1)^p \binom{p}{q}$$

$$= (n-n')\frac{1}{n^2} \sum_{p=0}^{n-n'-1} \binom{n-n'-1}{p}$$

$$\times \sum_{q\ge 0} (-1)^p \binom{p}{q} \sum_{m\ge q} \frac{1}{n^m}(m+1)! \frac{(n-1-p)^{m-q}}{(m-q)!}.$$

Simplifying the inner sum we get

$$\frac{1}{n^q} \sum_{m\ge 0} \frac{1}{n^m}(m+q+1)! \frac{(n-1-p)^m}{m!}$$

$$= \frac{(q+1)!}{n^q} \sum_{m\ge 0} \frac{1}{n^m} \binom{m+q+1}{q+1} (n-1-p)^m$$

$$= \frac{(q+1)!}{n^q} \frac{1}{(1-(n-1-p)/n)^{q+2}} = (q+1)! n^2 \frac{1}{(p+1)^{q+2}}.$$

We thus obtain for the sum of the probabilities

$$\sum_{m\ge 2} P[T=m] = (n-n') \sum_{p=0}^{n-n'-1} \binom{n-n'-1}{p} (-1)^p \sum_{q=0}^{p} \binom{p}{q} (q+1)! \frac{1}{(p+1)^{q+2}}.$$

Repeat to instantly obtain for **the expectation**

$$E[T] = n(n-n') \sum_{p=0}^{n-n'-1} \binom{n-n'-1}{p} (-1)^p \sum_{q=0}^{p} \binom{p}{q} \frac{(q+2)!}{(p+1)^{q+3}}.$$

Now to simplify these we start with the inner sum from the probablity using the fact that

$$\sum_{q=0}^{p} \binom{p}{q} (q+1)! \frac{1}{(p+1)^{q+1}} = 1$$

which was proved by residues at the cited link from the introduction. We then obtain

$$(n-n') \sum_{p=0}^{n-n'-1} \binom{n-n'-1}{p} \frac{(-1)^p}{p+1}$$

$$= \sum_{p=0}^{n-n'-1} \binom{n-n'}{p+1} (-1)^p = - \sum_{p=1}^{n-n'} \binom{n-n'}{p} (-1)^p$$

$$= 1 - \sum_{p=0}^{n-n'} {n-n' \choose p} (-1)^p = 1 - (1-1)^{n-n'} = 1$$

which confirms it being a probability distribution. We will not attempt this manipulation with the expectation, since actual computation of the values indicates that it does not simplify as announced earlier. For example, these are the expectations for the pairs $(2n', n')$:

$$4, 11, \frac{347}{18}, \frac{12259}{432}, \frac{41129339}{1080000}, \frac{390968681}{8100000}, \frac{336486120012803}{5717741400000}, \ldots$$

and for pairs $(3n', n')$:

$$\frac{33}{4}, \frac{12259}{576}, \frac{390968681}{10800000}, \frac{2859481756726972261}{54646360473600000}, \ldots$$

The reader who seeks numerical evidence confirming the closed form or additional clarification of the problem definition used is asked to consult the following simple C program whose output matched the formula on all cases that were examined.

This was math.stackexchange.com problem 2720594.

4 Drawing coupons until at least one instance of each type is seen with n' types already collected

Using the notation from the previous section we have from first principles that

$$\mathrm{P}[T = m] = \frac{1}{n^m} \times {n-n' \choose 1} \times (m-1)! [z^{m-1}] \exp(n'z) (\exp(z) - 1)^{n-n'-1}.$$

We shall see that with this closed form for the probabilities, we can not only compute the expectation of the number of draws to collect the remaining coupons but also the second factorial moment if desired, and the variance. To start **verify that this is a probability distribution**. We get

$$(n-n') \sum_{m \ge n-n'} \frac{1}{n^m} \times (m-1)! [z^{m-1}] \exp(n'z) (\exp(z) - 1)^{n-n'-1}$$

$$= (n-n') \sum_{m \ge n-n'} \frac{1}{n^m} (m-1)! [z^{m-1}] \exp(n'z)$$

$$\times \sum_{q=0}^{n-n'-1} {n-n'-1 \choose q} (-1)^{n-n'-1-q} \exp(qz)$$

$$= (n-n') \sum_{m\ge n-n'} \frac{1}{n^m} \sum_{q=0}^{n-n'-1} \binom{n-n'-1}{q} (-1)^{n-n'-1-q} (n'+q)^{m-1}$$

$$= \frac{n-n'}{n} \sum_{q=0}^{n-n'-1} \binom{n-n'-1}{q} (-1)^{n-n'-1-q} \sum_{m\ge n-n'} \frac{1}{n^{m-1}} (n'+q)^{m-1}$$

$$= \frac{n-n'}{n} \sum_{q=0}^{n-n'-1} \binom{n-n'-1}{q} (-1)^{n-n'-1-q} \frac{(n'+q)^{n-n'-1}/n^{n-n'-1}}{1-(n'+q)/n}$$

$$= \frac{n-n'}{n^{n-n'-1}} \sum_{q=0}^{n-n'-1} \binom{n-n'-1}{n-n'-1-q} (-1)^{n-n'-1-q} \frac{(n'+q)^{n-n'-1}}{n-n'-q}$$

$$= \frac{1}{n^{n-n'-1}} \sum_{q=0}^{n-n'-1} \binom{n-n'}{n-n'-q} (-1)^{n-n'-1-q} (n'+q)^{n-n'-1}$$

$$= -\frac{1}{n^{n-n'-1}} \sum_{q=0}^{n-n'-1} \binom{n-n'}{q} (-1)^{n-n'-q} (n'+q)^{n-n'-1}$$

$$= 1 - \frac{1}{n^{n-n'-1}} \sum_{q=0}^{n-n'} \binom{n-n'}{q} (-1)^{n-n'-q} (n'+q)^{n-n'-1}$$

$$= 1 - (n-n'-1)! [z^{n-n'-1}] \frac{\exp(n'z)}{n^{n-n'-1}} \sum_{q=0}^{n-n'} \binom{n-n'}{q} (-1)^{n-n'-q} \exp(qz)$$

$$= 1 - (n-n'-1)! [z^{n-n'-1}] \frac{\exp(n'z)}{n^{n-n'-1}} (\exp(z)-1)^{n-n'}.$$

Note however that $\exp(z) - 1 = z + \cdots$ and hence $(\exp(z)-1)^{n-n'} = z^{n-n'} + \cdots$ which means the coefficient extractor $[z^{n-n'-1}]$ is zero and we are left with just the first term, which is one, and we indeed have a probability distribution.

Continuing with the expectation we evidently require

$$\sum_{m\ge n-n'} \frac{m}{n^{m-1}} (n'+q)^{m-1}$$

$$= \frac{(n'+q)^{n-n'-1}}{n^{n-n'-1}} \sum_{m\ge 1} \frac{m+n-n'-1}{n^{m-1}} (n'+q)^{m-1}.$$

The simple component from this is

$$(n-n'-1) \frac{(n'+q)^{n-n'-1}}{n^{n-n'-1}} \frac{1}{1-(n'+q)/n}.$$

Here we recognize a term that we have already evaluated which yields on substitution into the outer sum the value $n-n'-1$. Evaluating the second term we get for the expectation

$$n-n'-1-\frac{1}{n^{n-n'-1}}\sum_{q=0}^{n-n'-1}\binom{n-n'}{q}(-1)^{n-n'-q}\frac{(n'+q)^{n-n'-1}}{1-(n'+q)/n}$$

or

$$\mathrm{E}[T]=n-n'-1-\frac{1}{n^{n-n'-2}}\sum_{q=0}^{n-n'-1}\binom{n-n'}{q}(-1)^{n-n'-q}\frac{(n'+q)^{n-n'-1}}{n-n'-q}.$$

Introducing

$$f(z)=\frac{(n-n')!}{n-n'-z}(n'+z)^{n-n'-1}\prod_{p=0}^{n-n'}\frac{1}{z-p}$$

we observe that for $0\le q\le n-n'-1$

$$\begin{aligned}\mathrm{Res}_{z=q}f(z)&=\frac{(n-n')!}{n-n'-q}(n'+q)^{n-n'-1}\prod_{p=0}^{q-1}\frac{1}{q-p}\prod_{p=q+1}^{n-n'}\frac{1}{q-p}\\&=\frac{(n-n')!}{n-n'-q}(n'+q)^{n-n'-1}\frac{1}{q!}\frac{(-1)^{n-n'-q}}{(n-n'-q)!}\end{aligned}$$

so that the expectation becomes

$$n-n'-1-\frac{1}{n^{n-n'-2}}\sum_{q=0}^{n-n'-1}\mathrm{Res}_{z=q}f(z).$$

Now residues sum to zero and the residue at infinity is zero as well since $\lim_{R\to\infty}2\pi R\times R^{n-n'-1}/R/R^{n-n'+1}=0$. So the sum is minus the residue at $z=n-n'$:

$$\mathrm{Res}_{z=n-n'}\frac{(n-n')!}{z-(n-n')}(n'+z)^{n-n'-1}\prod_{p=0}^{n-n'}\frac{1}{z-p}.$$

This needs

$$(n-n')!\left((n'+z)^{n-n'-1}\prod_{p=0}^{n-n'-1}\frac{1}{z-p}\right)'\Bigg|_{z=n-n'}$$

Note that when we are waiting for one last coupon i.e. $n = n' + 1$ the sum formula yields for the expectation $0 - n \times (-1) = n$ so we may suppose that $n > n' + 1$. Continue with the derivative to get

$$(n-n')!\; (n-n'-1)(n'+z)^{n-n'-2} \prod_{p=0}^{n-n'-1} \frac{1}{z-p} \Bigg|_{z=n-n'}$$

$$-(n-n')!\; (n'+z)^{n-n'-1} \prod_{p=0}^{n-n'-1} \frac{1}{z-p} \sum_{p=0}^{n-n'-1} \frac{1}{z-p} \Bigg|_{z=n-n'}$$

$$= (n-n'-1)n^{n-n'-2} - n^{n-n'-1} H_{n-n'}.$$

Replacing this in the main formula yields the closed form (which also produces the correct value for $n - n' = 1$ BTW)

$$\mathrm{E}[T] = n \times H_{n-n'} \quad \sim \quad n \log(n-n') + \gamma n + \frac{1}{2}\frac{n}{n-n'} - \frac{1}{12}\frac{n}{(n-n')^2} + \cdots .$$

We thus obtain for forty coupons with thirty already seen the expectation

$$\frac{7381}{63} \approx 117.1587302.$$

Moving on to conclude with the variance we now work with

$$\sum_{m \ge n-n'} \frac{m^2}{n^{m-1}} (n'+q)^{m-1}$$

$$= \frac{(n'+q)^{n-n'-1}}{n^{n-n'-1}} \sum_{m \ge 1} \frac{(m+n-n'-1)^2}{n^{m-1}} (n'+q)^{m-1}.$$

Here we recognize two easy pieces which are

$$(n-n'-1)^2$$

and

$$2(n-n'-1)(nH_{n-n'} - (n-n'-1)).$$

With $\sum_{m \ge 1} m^2 w^{m-1} = (1+w)/(1-w)^3$ we have two additional sum terms:

$$-\frac{1}{n^{n-n'-3}} \sum_{q=0}^{n-n'-1} \binom{n-n'}{q} (-1)^{n-n'-q} \frac{(n'+q)^{n-n'-1}}{(n-n'-q)^2}$$

and

$$-\frac{1}{n^{n-n'-2}} \sum_{q=0}^{n-n'-1} \binom{n-n'}{q} (-1)^{n-n'-q} \frac{(n'+q)^{n-n'}}{(n-n'-q)^2}.$$

For the first of these we use $f(z)/(n-n'-z)$ and obtain five pieces:

$$(n-n'-1)(n-n'-2)n^{n-n'-3} - 2(n-n'-1)n^{n-n'-2}H_{n-n'}$$
$$+n^{n-n'-1}H_{n-n'}^2 + n^{n-n'-1}H_{n-n'}^{(2)}.$$

The second sum only differs in the exponent on $n'+q$ and we obtain

$$(n-n')(n-n'-1)n^{n-n'-2} - 2(n-n')n^{n-n'-1}H_{n-n'}$$
$$+n^{n-n'}H_{n-n'}^2 + n^{n-n'}H_{n-n'}^{(2)}.$$

Collecting everything including a factor of $1/2$ on the derivative we finally have (observe cancelation of the polynomial in n and n')

$$\mathrm{E}[T^2] = n^2 \times H_{n-n'}^2 - n \times H_{n-n'} + n^2 \times H_{n-n'}^{(2)}.$$

Using that

$$\mathrm{Var}[T] = \mathrm{E}[T^2] - \mathrm{E}[T]^2$$

we get

$$\mathrm{Var}[T] = n^2 \times H_{n-n'}^{(2)} - n \times H_{n-n'}.$$

The dominant term here is $\sim \frac{\pi^2}{6} n^2$.

These results for the expectation and the variance are in agreement with Wikipedia on the coupon collector problem, where they are derived by probabilistic methods as opposed to the Stirling numbers used here.

This was math.stackexchange.com problem 2824168.

5 Processing until a multiset containing some number of distinct elements has been seen

We present the complexity using Stirling numbers of the second kind. Using the notation from this MSE link we have n coupons, and ask about the expected time until a multiset containing instances of j different coupons has been drawn.

First let us verify that we indeed have a probability distribution here. We have for the number T of coupons being m draws that

$$P[T=m] = \frac{1}{n^m} \times \binom{n}{j-1} \times \left\{ {m-1 \atop j-1} \right\} \times (j-1)! \times (n+1-j).$$

What happens here is that for a run of m samples to produce a multiset containing instances of j different coupons for the first time on the last sample we have two parts, a prefix of length $m-1$ and a terminal sample that completes the set. Therefore we must choose the $j-1$ values excluding the one that occurs last for the prefix from the n possibilities which gives the first binomial coefficient. Next we partition the first $m-1$ slots into $j-1$ non-empty sets in an ordered set partition. (Stirling number and factorial). The smallest value chosen gets the slots listed in the first set, the next one those in the second set etc. Finally we get $n-(j-1)$ possibilities ($j-1$ values from the prefix have already been used) for the terminal sample that completes the selection. Combine with n^m possible choices.

Recall the OGF of the Stirling numbers of the second kind which says that

$$\left\{ {n \atop k} \right\} = [z^n] \prod_{q=1}^{k} \frac{z}{1-qz}.$$

This gives for the sum of the probabilities

$$\sum_{m\ge 1} P[T=m] = {n \choose j-1} (j-1)!(n+1-j) \frac{1}{n} \sum_{m\ge 1} \frac{1}{n^{m-1}} \left\{ {m-1 \atop j-1} \right\}.$$

Focusing on the sum we obtain

$$\sum_{m\ge 1} \frac{1}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{j-1} \frac{z}{1-qz} = \prod_{q=1}^{j-1} \frac{1/n}{1-q/n}$$
$$= \prod_{q=1}^{j-1} \frac{1}{n-q} = \frac{(n-j)!}{(n-1)!}.$$

Combining this with the outer factor we get

$${n \choose j-1} (j-1)!(n+1-j) \frac{1}{n} \frac{(n-j)!}{(n-1)!}$$
$$= {n \choose j-1} (j-1)! \frac{(n+1-j)!}{n!} = 1$$

This confirms it being a probability distribution.

We then get for the expectation that

$$\sum_{m\ge 1} m \times P[T=m]$$
$$= {n \choose j-1} (j-1)!(n+1-j) \frac{1}{n} \sum_{m\ge 1} \frac{m}{n^{m-1}} \left\{ {m-1 \atop j-1} \right\}.$$

We once more focus on the sum to get

$$\sum_{m\ge 1}\frac{m}{n^{m-1}}[z^{m-1}]\prod_{q=1}^{j-1}\frac{z}{1-qz} = \sum_{m\ge 1}\frac{m}{n^{m-1}}[z^m]z\prod_{q=1}^{j-1}\frac{z}{1-qz}$$

$$= \left.\left(\prod_{q=0}^{j-1}\frac{z}{1-qz}\right)'\right|_{z=1/n}$$

$$= \left.\left(\prod_{q=0}^{j-1}\frac{z}{1-qz}\sum_{p=0}^{j-1}\frac{1-pz}{z}\frac{1}{(1-pz)^2}\right)\right|_{z=1/n}$$

$$= \left.\left(\prod_{q=0}^{j-1}\frac{z}{1-qz}\sum_{p=0}^{j-1}\frac{1}{z}\frac{1}{1-pz}\right)\right|_{z=1/n}$$

$$= \prod_{q=0}^{j-1}\frac{1/n}{1-q/n}\sum_{p=0}^{j-1}\frac{1}{1/n}\frac{1}{1-p/n}$$

$$= \prod_{q=0}^{j-1}\frac{1}{n-q}\sum_{p=0}^{j-1}\frac{n^2}{n-p} = n\prod_{q=1}^{j-1}\frac{1}{n-q}\sum_{p=0}^{j-1}\frac{1}{n-p}.$$

Retrieving the outer factor we have

$$\binom{n}{j-1}(j-1)!(n+1-j)\frac{1}{n}\frac{(n-j)!}{(n-1)!}\times n\sum_{p=0}^{j-1}\frac{1}{n-p}.$$

The front simplifies to one as before and we are left with

$$n\sum_{p=0}^{j-1}\frac{1}{n-p} = n\left(\sum_{p=0}^{n-1}\frac{1}{n-p} - \sum_{p=j}^{n-1}\frac{1}{n-p}\right).$$

This is

$$n\times(H_n - H_{n-j})$$

This yields nH_n when $j=n$ and 1 when $j=1$ which are both correct. Using $H_n \sim \log n + \gamma$ we get for $j=n/2$ the expectation $n\log 2$.

This was math.stackexchange.com problem 2021884.

6 Expected number of singletons once all coupons have been collected

Here is the expectation using Stirling numbers of the second kind. In referencing the notation from this MSE link we have n coupons, and ask about the expected number of singletons once a complete set of n different coupons has been drawn. We will be using OGFs and EGFs of Stirling numbers and switch between them.

First let us verify that we indeed have a probability distribution here. We have for the number T of coupons being m draws classified according to the number of singletons that

$$P[T=m] = \frac{1}{n^m} \times \binom{n}{n-1}$$
$$\times \sum_{q=0}^{n-1} \binom{n-1}{q} \binom{m-1}{q} q! \left\{ m-1-q \atop n-1-q \right\}_{\ge 2} (n-1-q)!.$$

What is happening here is that we first choose the $n-1$ types of coupons that go into the prefix, where the one not selected goes into the suffix, completing the set of coupons. Next we choose q coupons from the ones in the prefix which will be represented by singletons (factor $\binom{n-1}{q}$). Next we choose the positions from the available slots where the singletons will be placed (factor $\binom{m-1}{q}q!$). We split the leftover $m-1-q$ slots into sets of at least two elements, one for each of the $n-1-q$ types that have not been instantiated (factor $\left\{ m-1-q \atop n-1-q \right\}_{\ge 2}(n-1-q)!$).

This probability simplifies to

$$P[T=m] = \frac{n \times (m-1)!}{n^m} \sum_{q=0}^{n-1} \frac{(n-1)!}{q!} \frac{1}{(m-1-q)!} \left\{ m-1-q \atop n-1-q \right\}_{\ge 2}$$
$$= \frac{n \times (m-1)!}{n^m} \sum_{q=0}^{n-1} \frac{(n-1)!}{q!} [z^{m-1-q}] \frac{(\exp(z)-z-1)^{n-1-q}}{(n-1-q)!}$$
$$= \frac{n \times (m-1)!}{n^m} \sum_{q=0}^{n-1} \binom{n-1}{q} [z^{m-1-q}] (\exp(z)-z-1)^{n-1-q}$$
$$= \frac{n \times (m-1)!}{n^m} \sum_{q=0}^{n-1} \binom{n-1}{q} [z^{m-1}] z^q (\exp(z)-z-1)^{n-1-q}$$
$$= \frac{n \times (m-1)!}{n^m} [z^{m-1}] (\exp(z)-1)^{n-1}$$
$$= \frac{n! \times (m-1)!}{n^m} [z^{m-1}] \frac{(\exp(z)-1)^{n-1}}{(n-1)!}.$$

We then get for the sum of the probabilities (observe that the EGF has morphed into an OGF)

$$\sum_{m\ge 1} P[T=m] = \frac{n!}{n} \sum_{m\ge 1} \frac{1}{n^{m-1}} [z^{m-1}] \prod_{q=1}^{n-1} \frac{z}{1-qz} = \frac{n!}{n} \prod_{q=1}^{n-1} \frac{1/n}{1-q/n}$$
$$= \frac{n!}{n} \prod_{q=1}^{n-1} \frac{1}{n-q} = \frac{n!}{n} \frac{1}{(n-1)!} = 1.$$

The probabilities sum to one and the sanity check goes through.

Continuing with the expected number of singletons we get an extra factor q which yields

$$\frac{n \times (m-1)!}{n^m} \sum_{q=1}^{n-1} q \binom{n-1}{q} [z^{m-1}] z^q (\exp(z) - z - 1)^{n-1-q}$$

$$= \frac{n(n-1) \times (m-1)!}{n^m} [z^{m-1}] \sum_{q=1}^{n-1} \binom{n-2}{q-1} z^q (\exp(z) - z - 1)^{n-1-q}$$

$$= \frac{n(n-1) \times (m-1)!}{n^m}$$

$$\times [z^{m-1}] z \sum_{q=1}^{n-1} \binom{n-2}{q-1} z^{q-1} (\exp(z) - z - 1)^{n-2-(q-1)}$$

$$= \frac{n(n-1) \times (m-1)!}{n^m} [z^{m-2}] (\exp(z) - 1)^{n-2}$$

$$= \frac{n! \times (m-1)!}{n^m} [z^{m-2}] \frac{(\exp(z) - 1)^{n-2}}{(n-2)!}.$$

Now we have

$$\sum_{m \ge 2} w^{m-2} (m-1)! [z^{m-2}] \sum_{q \ge 0} A_q \frac{z^q}{q!}$$

$$= \sum_{m \ge 2} w^{m-2} (m-1) A_{m-2} = \left. \left(z \sum_{q \ge 0} A_q z^q \right)' \right|_{z=w}.$$

Applying this to the expectation yields

$$\frac{n!}{n^2} \sum_{m \ge 2} \frac{1}{n^{m-2}} [z^{m-2}] \left(\prod_{q=0}^{n-2} \frac{z}{1-qz} \right)'$$

$$= \frac{n!}{n^2} \sum_{m \ge 2} \frac{1}{n^{m-2}} [z^{m-2}] \prod_{q=0}^{n-2} \frac{z}{1-qz} \sum_{q=0}^{n-2} \frac{1-qz}{z} \frac{1}{(1-qz)^2}$$

$$= \frac{n!}{n^2} \sum_{m \ge 2} \frac{1}{n^{m-2}} [z^{m-2}] \prod_{q=0}^{n-2} \frac{z}{1-qz} \sum_{q=0}^{n-2} \frac{1/z}{1-qz}$$

$$= \frac{n!}{n^2} \prod_{q=0}^{n-2} \frac{1/n}{1-q/n} \sum_{q=0}^{n-2} \frac{n}{1-q/n}$$

$$= n! \prod_{q=0}^{n-2} \frac{1}{n-q} \sum_{q=0}^{n-2} \frac{1}{n-q}.$$

This simplifies to the end result

$$H_n \sim \log n + \gamma$$

where we have included an increment of one that represents the singleton which completed the set of coupons.

This post made extensive use of the technique of *annihilated coefficient extractors* (ACE). There are more of these at this MSE link I and at this MSE link II and also here at this MSE link III.

This was math.stackexchange.com problem 2045183.

6.1 The coupon collector's sibling

This problem is closely related and represents the scenario where the coupon collector gives his duplicates to his sibling once all coupons have been collected and we ask how many coupons the sibling is missing for a full collection on average.

We start with the species of ordered set partitions with sets of more than two elements marked. This is

$$\text{SEQ}(\mathcal{UZ} + \mathcal{UV}\text{SET}_{\ge 2}(\mathcal{Z})).$$

We thus obtain the generating function

$$G(z, u, v) = \frac{1}{1 - u(v \exp(z) - vz + z - 1)}.$$

We then get for the probability that

$$P[T = m] = \frac{1}{n^m} \binom{n}{n-1} (m-1)! [z^{m-1}][u^{n-1}] G(z, u, v).$$

What happens here is very simple. We choose the $n-1$ coupons that go into the prefix consisting of $m-1$ draws. Then we partition those draws into sets, one for each type of coupon, containing the position where it appeared. We mark sets of more than two elements. Doing the extraction in u we find

$$P[T = m] = \frac{1}{n^m} \binom{n}{n-1} (m-1)! [z^{m-1}] (v \exp(z) - vz + z - 1)^{n-1}.$$

Now to do the usual sanity check that we have a probability distribution we remove the marking in v and obtain

$$\sum_{m \ge 1} P[T = m] = \sum_{m \ge 1} \frac{n!}{n^m} (m-1)! [z^{m-1}] \frac{(\exp(z) - 1)^{n-1}}{(n-1)!}.$$

This was evaluated at the cited link and the sanity check goes through, more or less by inspection in fact. Continuing with the expectation of coupons that were drawn mor than once we differentiate with respect to v and set $v = 1$, getting

$$\left.\frac{n! \times (m-1)!}{n^m}[z^{m-1}](n-1)\; \frac{(v\exp(z) - vz + z - 1)^{n-2}}{(n-1)!} \times (\exp(z) - z)\right|_{v=1}$$

$$= \frac{n! \times (m-1)!}{n^m}[z^{m-1}]\frac{(\exp(z)-1)^{n-2}}{(n-2)!} \times (\exp(z) - z).$$

We write this in three pieces, namely

$$\frac{n! \times (m-1)!}{n^m}[z^{m-1}]\frac{(\exp(z)-1)^{n-1}}{(n-2)!}$$
$$-\frac{n! \times (m-1)!}{n^m}[z^{m-2}]\frac{(\exp(z)-1)^{n-2}}{(n-2)!}$$
$$+\frac{n! \times (m-1)!}{n^m}[z^{m-1}]\frac{(\exp(z)-1)^{n-2}}{(n-2)!}.$$

Consulting the results from the main link we find for the first two pieces

$$n - 1 - (H_n - 1) = n - H_n.$$

We then get for the third piece (recognizing the Stirling number EGF and observing that the EGF morphs into an OGF)

$$\frac{n!}{n}\sum_{m\ge 1}\frac{1}{n^{m-1}}[z^{m-1}]\prod_{q=1}^{n-2}\frac{z}{1-qz} = \frac{n!}{n}\prod_{q=1}^{n-2}\frac{1/n}{1-q/n}$$

$$= \frac{n!}{n}\prod_{q=1}^{n-2}\frac{1}{n-q} = \frac{n!}{n}\frac{1}{(n-1)!} = 1.$$

We thus have for the answer that the sibling collects $n + 1 - H_n$ coupons and hence is missing $H_n - 1$ coupons probabilistically from among the coupons collected in the prefix. Furthermore and deterministically, the sibling never sees the last coupon collected because it is always a singleton. Hence the sibling is missing

$$H_n$$

coupons. We may add the halting singleton because it does not involve any additional probability and is determined by the set partition of the prefix.

What have we learned? On seeing this result it immediately becomes evident that these two parameters (singletons and duplicates) are prefectly additive on the level of generating functions and we could have concluded by inspection, citing the result for singletons from the link without any extra calculation.

This was math.stackexchange.com problem 2166203.

7 Computing the expectation of T-choose-Q, with T the number of steps and Q the number of distinct coupons among the first j retrieved

We present a problem inspired by the work at this MSE link. In particular, we consider a coupon collector scenario with n coupons where an integer $1 \le j \le n-1$ is given. We introduce two random variables, namely T and Q where T represents the number of draws until all coupons have been collected and Q the number of different coupons that appeared in the first j draws. The following conjecture is submitted for your consideration.

$$\mathrm{E}\left[\binom{T}{Q}\right] = \sum_{k=1}^{j} \frac{n!}{n^{n-k-1+j}} \times \left\{ j \atop k \right\} \sum_{r=0}^{k} \binom{n+j-k}{k-r}$$
$$\times \sum_{p=0}^{n-k-1} \frac{(-1)^{n-k-1-p}}{p!(n-k-1-p)!} \frac{(k+p)^{n-k-1+r}}{(n-k-p)^{r+1}}.$$

I have what I believe to be a proof but it is quite involved. We propose the following list of questions concerning the above identity:

- does it indeed hold and does it perhaps have a straightforward proof using probabilistic methods and is there structural simplification
- what are the asymptotics, are there effective estimates of these terms that match the numeric exact values from the formula without having recurse to a triple sum.

The reader is invited to compare potentially relevant asymptotics to the data from the identity.

There is the following extremely basic C program which I include here to help clarify what interpretation of the problem is being used. Compiled with GCC 4.3.2 and the **std=gnu99** option.

Addendum. As a sanity check when $j = 1$ the formula should produce nH_n for $n \ge 2$. In fact we obtain

$$\frac{n!}{n^{n-1}} \left(n \times \sum_{p=0}^{n-2} \frac{(-1)^{n-2-p}}{p!(n-2-p)!} \frac{(1+p)^{n-2}}{n-1-p} + \sum_{p=0}^{n-2} \frac{(-1)^{n-2-p}}{p!(n-2-p)!} \frac{(1+p)^{n-1}}{(n-1-p)^2} \right).$$

For the first sum we introduce

$$f(z) = \frac{(1+z)^{n-2}}{n-1-z} \prod_{q=0}^{n-2} \frac{1}{z-q}$$

so that the sum is given by (residues sum to zero)

$$\sum_{q=0}^{n-2} \mathrm{Res}_{z=q} f(z) = -\mathrm{Res}_{z=n-1} f(z) - \mathrm{Res}_{z=\infty} f(z).$$

The contribution from the first term is

$$\frac{n^{n-2}}{(n-1)!}$$

and from the second

$$\mathrm{Res}_{z=0} \frac{1}{z^2} \frac{(1+1/z)^{n-2}}{n-1-1/z} \prod_{q=0}^{n-2} \frac{1}{1/z-q} = \mathrm{Res}_{z=0} \frac{1}{z^n} \frac{(1+z)^{n-2}}{n-1-1/z} \prod_{q=0}^{n-2} \frac{z}{1-qz}$$

$$= \mathrm{Res}_{z=0} \frac{1}{z} \frac{(1+z)^{n-2}}{n-1-1/z} \prod_{q=0}^{n-2} \frac{1}{1-qz} = \mathrm{Res}_{z=0} \frac{(1+z)^{n-2}}{z(n-1)-1} \prod_{q=0}^{n-2} \frac{1}{1-qz} = 0.$$

Hence the first sum contributes

$$\frac{n!}{n^{n-1}} \times n \frac{n^{n-2}}{(n-1)!} = n.$$

For the second sum we use

$$g(z) = \frac{(1+z)^{n-1}}{(n-1-z)^2} \prod_{q=0}^{n-2} \frac{1}{z-q} = \frac{(1+z)^{n-1}}{(z-(n-1))^2} \prod_{q=0}^{n-2} \frac{1}{z-q}.$$

We get for the negative of the residue at $n-1$ the value

$$-\left((1+z)^{n-1} \prod_{q=0}^{n-2} \frac{1}{z-q} \right)'_{z=n-1}$$

$$= -\left((n-1)(1+z)^{n-2} \prod_{q=0}^{n-2} \frac{1}{z-q} - (1+z)^{n-1} \prod_{q=0}^{n-2} \frac{1}{z-q} \sum_{q=0}^{n-2} \frac{1}{z-q} \right)_{z=n-1}$$

$$= -\left((n-1)n^{n-2} \frac{1}{(n-1)!} - n^{n-1} \frac{1}{(n-1)!} H_{n-1} \right).$$

Multiply by $n!/n^{n-1}$ to get

$$nH_{n-1} - (n-1)n^{n-2} \frac{1}{(n-1)!} \frac{n!}{n^{n-1}}$$

$$= nH_{n-1} - (n-1)\frac{n}{n} = nH_{n-1} - (n-1).$$

For the negative of the residue at infinity we obtain

$$\mathrm{Res}_{z=0}\frac{1}{z^2}\frac{(1+1/z)^{n-1}}{(n-1-1/z)^2}\prod_{q=0}^{n-2}\frac{1}{1/z-q} = \mathrm{Res}_{z=0}\frac{1}{z^{n+1}}\frac{(1+z)^{n-1}}{(n-1-1/z)^2}\prod_{q=0}^{n-2}\frac{z}{1-qz}$$

$$= \mathrm{Res}_{z=0}\frac{1}{z^2}\frac{(1+z)^{n-1}}{(n-1-1/z)^2}\prod_{q=0}^{n-2}\frac{1}{1-qz}$$

$$= \mathrm{Res}_{z=0}\frac{(1+z)^{n-1}}{(z(n-1)-1)^2}\prod_{q=0}^{n-2}\frac{1}{1-qz} = 0.$$

Collecting everything we get

$$nH_{n-1} - (n-1) + n = nH_{n-1} + n\frac{1}{n}$$

or alternatively

$$nH_n$$

and the sanity check goes through. Observe that we evidently require something more sophisticated to prove the conjectured identity e.g. when $j = n-1$. (Remark. We don't need to actually apply the formula for the residues at infinity, it is sufficient when working with rational functions to observe that both $f(z)$ and $g(z)$ have the difference between the degree of the denominator and of the numerator equal to two.)

This was math.stackexchange.com problem 2125064.

7.1 Proof

We use the notation from the following MSE link with m for the number of rolls and n for the number of coupons. We can actually answer a more general question, namely what is the expected number of different faces in the first j rolls where $j \le n-1$. We classify according to the number k of different faces that appeared where $1 \le k \le j$. There are at least two types of coupons.

First let us verify that we indeed have a probability distribution here. We have for the number T of coupons being m draws that the number of configurations i.e. admissible sequences of draws is

$$\binom{n}{k} \times \left\{ {j \atop k} \right\} \times k! \times (n-k)$$

$$\times \sum_{p=0}^{k} \binom{k}{p} \left\{ {m-1-j \atop p+n-k-1} \right\} \times (p+n-k-1)!.$$

Observe that

$$\sum_{p=0}^{k}\binom{k}{p}\left\{ m-1-j \atop p+n-k-1 \right\} \times (p+n-k-1)!$$

$$= (m-1-j)![z^{m-1-j}]\sum_{p=0}^{k}\binom{k}{p}(\exp(z)-1)^{p+n-k-1}$$

$$= (m-1-j)![z^{m-1-j}](\exp(z)-1)^{n-k-1}\sum_{p=0}^{k}\binom{k}{p}(\exp(z)-1)^{p}$$

$$= (m-1-j)![z^{m-1-j}](\exp(z)-1)^{n-k-1}\exp(kz).$$

This is

$$(n-k-1)!\sum_{p=0}^{m-1-j}\binom{m-1-j}{p}\left\{ p \atop n-k-1 \right\} \times k^{m-1-j-p}.$$

We thus introduce the generating function

$$G_{j,k}(z) = \sum_{m\ge n} z^m \sum_{p=0}^{m-1-j}\binom{m-1-j}{p}\left\{ p \atop n-k-1 \right\} \times k^{m-1-j-p}.$$

Now put

$$\binom{m-1-j}{p} = \frac{1}{2\pi i}\int_{|w|=\epsilon}\frac{1}{w^{m-j-p}}\frac{1}{(1-w)^{p+1}}\; dw$$

which controls the range so we may extend p to infinity which yields

$$G_{j,k}(z) = k^{-1-j}\sum_{m\ge n} z^m k^m$$

$$\times\frac{1}{2\pi i}\int_{|w|=\epsilon}\frac{1}{w^{m-j}}\frac{1}{1-w}\sum_{p\ge 0}\left\{ p \atop n-k-1 \right\}\frac{w^p}{(1-w)^p} \times k^{-p}\; dw.$$

Recall the OGF of the Stirling numbers of the second kind which says that

$$\left\{ n \atop k \right\} = [z^n]\prod_{q=1}^{k}\frac{z}{1-qz}.$$

In the present case this yields

$$G_{j,k}(z) = k^{-1-j}\sum_{m\ge n} z^m k^m$$

$$\times\frac{1}{2\pi i}\int_{|w|=\epsilon}\frac{1}{w^{m-j}}\frac{1}{1-w}\prod_{q=1}^{n-k-1}\frac{w/(1-w)/k}{1-qw/(1-w)/k}\; dw$$

$$= k^{-1-j} \sum_{m\ge n} z^m k^m$$

$$\times \frac{1}{2\pi i} \int_{|w|=\epsilon} \frac{1}{w^{m+1}} \frac{w^{j+1}}{1-w} \prod_{q=1}^{n-k-1} \frac{w/(1-w)/k}{1-qw/(1-w)/k} \; dw$$

$$= k^{-1-j} \frac{k^{j+1} z^{j+1}}{1-kz} \prod_{q=1}^{n-k-1} \frac{z/(1-kz)}{1-qz/(1-kz)}$$

$$= \frac{z^{j+1}}{1-kz} \prod_{q=1}^{n-k-1} \frac{z}{1-kz-qz} = \frac{z^{n+j-k}}{1-kz} \prod_{q=1}^{n-k-1} \frac{1}{1-kz-qz}$$

$$= z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz}.$$

We have shown that for the probability of having m draws we get

$$P[T=m] = \frac{n!}{n^m} \sum_{k=1}^{j} \left\{ {j \atop k} \right\} [z^m] z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz}.$$

This gives for the sum of the probabilities

$$n! \sum_{k=1}^{j} \left\{ {j \atop k} \right\} \frac{1}{n^{n+j-k}} \prod_{q=0}^{n-k-1} \frac{1}{1-k/n-q/n}$$

$$= n! \sum_{k=1}^{j} \left\{ {j \atop k} \right\} \frac{1}{n^{n+j-k}} \prod_{q=0}^{n-k-1} \frac{n}{n-k-q} = n! \sum_{k=1}^{j} \left\{ {j \atop k} \right\} \frac{1}{n^j} \prod_{q=0}^{n-k-1} \frac{1}{n-k-q}$$

$$= \frac{n!}{n^j} \sum_{k=1}^{j} \left\{ {j \atop k} \right\} \frac{1}{(n-k)!} = \frac{1}{n^j} \sum_{k=1}^{j} \binom{n}{k} k! \left\{ {j \atop k} \right\} = \frac{1}{n^j} j! [z^j] \sum_{k=1}^{j} \binom{n}{k} (\exp(z)-1)^k.$$

Now since $\exp(z)-1$ starts at z the power k starts at z^k. Therefore we may extend the range of k beyond j without adding any terms (coefficient on $[z^j]$ being extracted). We may also include $k=0$, which is a number. We obtain

$$\frac{1}{n^j} j! [z^j] \sum_{k=0}^{n} \binom{n}{k} (\exp(z)-1)^k = \frac{1}{n^j} j! [z^j] \exp(nz) = \frac{1}{n^j} j! \frac{n^j}{j!} = 1.$$

This confirms it being a probability distribution.

Moving on to the expectation we evidently require the following quantity:

$$\sum_{m\ge n} \binom{m}{k} \times \frac{n!}{n^m} \times \left\{ {j \atop k} \right\} [z^m] z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz}$$

$$= \frac{n!}{n^k} \times \left\{ {j \atop k} \right\} \times \sum_{m\ge n} \binom{m}{k} \frac{1}{n^{m-k}} [z^m] z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz}$$

$$= \frac{n!}{k! \times n^k} \times \left\{ {j \atop k} \right\} \times \sum_{m\ge n} \frac{1}{n^{m-k}} m^{\underline{k}} [z^m] z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz}$$

$$= \frac{n!}{k! \times n^k} \times \left\{ {j \atop k} \right\} \times \left. \left(z^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-kz-qz} \right)^{(k)} \right|_{z=1/n.}$$

We move to deploy the generalized Leibniz rule which requires

$$\sum_{p\ge 0} \frac{1}{p!} (z^{n+j-k})^{(p)} w^p = \sum_{p\ge 0} \binom{n+j-k}{p} z^{n+j-k-p} w^p$$

$$= z^{n+j-k} \left(1 + \frac{w}{z}\right)^{n+j-k} = (w+z)^{n+j-k}$$

as well as

$$\sum_{p\ge 0} \frac{1}{p!} \left(\frac{1}{1-kz-qz} \right)^{(p)} w^p = \sum_{p\ge 0} \frac{(k+q)^p}{(1-kz-qz)^{p+1}} w^p$$

$$= \frac{1}{1-kz-qz} \frac{1}{1-(k+q)w/(1-kz-qz)} = \frac{1}{1-(k+q)(w+z)}.$$

Hence the substituted derivative is

$$k![w^k](w+1/n)^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-(k+q)(w+1/n)}$$

which yields for the sum

$$\frac{n!}{n^k} \times \left\{ {j \atop k} \right\} \times [w^k](w+1/n)^{n+j-k} \prod_{q=0}^{n-k-1} \frac{1}{1-(k+q)(w+1/n)}.$$

Prepare for partial fractions by residues on the product term which yields

$$\prod_{q=0}^{n-k-1} \frac{1}{k+q} \prod_{q=0}^{n-k-1} \frac{1}{1/(k+q)-(w+1/n)}$$

$$= \frac{(k-1)!}{(n-1)!} (-1)^{n-k} \prod_{q=0}^{n-k-1} \frac{1}{w-(1/(k+q)-1/n)}.$$

We get for the residue at $w = 1/(k+p) - 1/n$

$$\prod_{q=0, q\ne p}^{n-k-1} \frac{1}{1/(k+p)-1/n-(1/(k+q)-1/n)}$$

$$= \prod_{q=0, q\neq p}^{n-k-1} \frac{1}{1/(k+p) - 1/(k+q)} = \prod_{q=0, q\neq p}^{n-k-1} \frac{(k+p)(k+q)}{q-p}$$

$$= \frac{(n-1)!}{(k-1)!} (-1)^p \frac{1}{p!} \frac{(k+p)^{n-k-2}}{(n-k-1-p)!}.$$

We thus obtain

$$\frac{n!}{n^k} \times \left\{ j \atop k \right\} \times [w^k](w+1/n)^{n+j-k}$$

$$\times \sum_{p=0}^{n-k-1} \frac{1}{w - (1/(k+p) - 1/n)} (-1)^{n-k+p} \frac{1}{p!} \frac{(k+p)^{n-k-2}}{(n-k-1-p)!}.$$

Observe that

$$[w^r] \frac{1}{w - (1/(k+p) - 1/n)}$$

$$= -\frac{1}{1/(k+p) - 1/n} [w^r] \frac{1}{1 - w/(1/(k+p) - 1/n)}$$

$$= -\frac{1}{(1/(k+p) - 1/n)^{r+1}}$$

and we obtain the sum form

$$-\frac{n!}{n^k} \times \left\{ j \atop k \right\} \sum_{r=0}^{k} [w^{k-r}](w+1/n)^{n+j-k}$$

$$\times \sum_{p=0}^{n-k-1} \frac{(-1)^{n-k-p}}{p!(n-k-1-p)!} (k+p)^{n-k-2} \frac{1}{(1/(k+p) - 1/n)^{r+1}}$$

$$= -\frac{n!}{n^k} \times \left\{ j \atop k \right\} \sum_{r=0}^{k} \binom{n+j-k}{k-r} \frac{1}{n^{n+j-2k+r}}$$

$$\times \sum_{p=0}^{n-k-1} \frac{(-1)^{n-k-p}}{p!(n-k-1-p)!} \frac{n^{r+1}(k+p)^{n-k-1+r}}{(n-k-p)^{r+1}}$$

$$= \frac{n!}{n^{n-k-1+j}} \times \left\{ j \atop k \right\} \sum_{r=0}^{k} \binom{n+j-k}{k-r}$$

$$\times \sum_{p=0}^{n-k-1} \frac{(-1)^{n-k-1-p}}{p!(n-k-1-p)!} \frac{(k+p)^{n-k-1+r}}{(n-k-p)^{r+1}}.$$

This was math.stackexchange.com problem 2125064.

8 Retrieving coupons in packets of unique coupons of a certain size

We can actually solve some special cases. Suppose we have n types of coupons which are drawn in packets of q coupons, with no duplicates in the packets. We derive closed forms for any packet size and evaluate them for $q = 2$. Now with $T_{m,n,q}$ the number of ways of drawing m packets of q-subsets of $[n]$ so that all possible values of n are present we get for the probability of m draws the closed form

$$\begin{aligned}P[T = m] &= \binom{n}{q}^{-m} \sum_{k=1}^{q} \binom{n}{n-k} \times T_{m-1,n-k,q} \times \binom{n-k}{q-k} \\ &= \binom{n}{q}^{-m} \sum_{k=1}^{q} \binom{n}{k} \times T_{m-1,n-k,q} \times \binom{n-k}{q-k}.\end{aligned}$$

This is for $n > q$ since the process always halts at the first step when $n = q$. The sum variable k is the count of the values missing before the draw of the subset at position m or alternatively of the values that appear in draw m for the first time.

To compute the terms in T we have a simple version of the computation at the following MSE link and introduce the generating function

$$[z^q] \prod_{l=1}^{n} (1 + zA_l)$$

which generates the q-subsets so that

$$\left([z^q] \prod_{l=1}^{n} (1 + zA_l) \right)^m$$

generates the m-sequences of q-subsets. We use inclusion-exclusion to remove those terms where some of the n terms are missing. The nodes $P \subseteq A$ in the poset represent terms from the generating function where the elements of P are missing plus possibly some more. This is evidently accomplished by setting the $A_l \in P$ to zero. We set the remaining A_l to one to obtain a count. The contribution for a given P is

$$[z^q](1 + z)^{n-|P|} = \binom{n - |P|}{q}.$$

Therefore inclusion-exclusion yields

$$\sum_{p=0}^{n} \binom{n}{p} (-1)^p \binom{n-p}{q}^m$$

which is

$$T_{m,n,q} = \sum_{p=0}^{n} \binom{n}{p} (-1)^{n-p} \binom{p}{q}^m .$$

This is zero when $m = 0$ and $n \ge 1$. Now as a sanity check we should have $T_{m,n,1} = {m \brace n} \times n!$ and indeed we obtain

$$\sum_{p=0}^{n} \binom{n}{p} (-1)^{n-p} \binom{p}{1}^m = \sum_{p=0}^{n} \binom{n}{p} (-1)^{n-p} p^m = {m \brace n} \times n!$$

and the check goes through. The rest is as shown at the following MSE link.

Next let us try to verify that we have a probability distribution. We have

$$\sum_{m\ge 1} P[T=m] = \sum_{k=1}^{q} \binom{n}{k} \binom{n-k}{q-k} \sum_{p=0}^{n-k} \binom{n-k}{p} (-1)^{n-k-p} \sum_{m\ge 1} \binom{n}{q}^{-m} \binom{p}{q}^{m-1}$$

$$= \sum_{k=1}^{q} \binom{n}{k} \binom{n-k}{q-k} \binom{n}{q}^{-1} \sum_{p=0}^{n-k} \binom{n-k}{p} (-1)^{n-k-p} \frac{1}{1 - \binom{p}{q}/\binom{n}{q}}$$

$$= \sum_{k=1}^{q} \binom{n}{k} \binom{n-k}{q-k} \sum_{p=0}^{n-k} \binom{n-k}{p} (-1)^{n-k-p} \left(\binom{n}{q} - \binom{p}{q} \right)^{-1}.$$

Specializing to $q = 2$ we get

$$2 \sum_{k=1}^{2} \binom{n}{k} \binom{n-k}{2-k} \sum_{p=0}^{n-k} \binom{n-k}{p} (-1)^{n-k-p} \frac{1}{n(n-1) - p(p-1)}$$

$$= 2n(n-1) \sum_{p=0}^{n-1} \binom{n-1}{p} (-1)^{n-1-p} \frac{1}{n(n-1) - p(p-1)}$$

$$+ n(n-1) \sum_{p=0}^{n-2} \binom{n-2}{p} (-1)^{n-2-p} \frac{1}{n(n-1) - p(p-1)}.$$

We evaluate the two sums by residues, using for the first sum

$$f(z) = \frac{(n-1)!}{n(n-1) - z(z-1)} \prod_{p=0}^{n-1} \frac{1}{z-p} = -\frac{(n-1)!}{(z-n)(z-(1-n))} \prod_{p=0}^{n-1} \frac{1}{z-p}.$$

The residue at infinity is zero and the residues at n and $1-n$ are

$$-\frac{(n-1)!}{2n-1} \frac{1}{n!} - \frac{(n-1)!}{1-2n} \frac{(-1)^n (n-2)!}{(2n-2)!}.$$

We get for the second sum by the same technique

$$-\frac{(n-2)!}{2n-1}\frac{1}{n!} - \frac{(n-2)!}{1-2n}\frac{(-1)^{n-1}(n-2)!}{(2n-3)!}.$$

Negate and add to get

$$n(n-1)\times$$

$$\left(\frac{1}{2n-1}\left(\frac{2}{n}+\frac{1}{n(n-1)}\right)+\frac{(n-2)!}{1-2n}\frac{(-1)^{n-1}(n-2)!}{(2n-3)!}\left(1-2\frac{n-1}{2n-2}\right)\right)$$

$$= \frac{1}{2n-1}(2n-2+1) = 1.$$

This confirms it being a probability distribution.
We now give a closed form for the expectation. We find

$$\sum_{m\ge 1} mP[T=m]$$

$$= \sum_{k=1}^{q}\binom{n}{k}\binom{n-k}{q-k}\sum_{p=0}^{n-k}\binom{n-k}{p}(-1)^{n-k-p}\sum_{m\ge 1} m\binom{n}{q}^{-m}\binom{p}{q}^{m-1}$$

$$= \sum_{k=1}^{q}\binom{n}{k}\binom{n-k}{q-k}\binom{n}{q}^{-1}\sum_{p=0}^{n-k}\binom{n-k}{p}(-1)^{n-k-p}\frac{1}{\left(1-\binom{p}{q}/\binom{n}{q}\right)^{2}}.$$

This is

$$\binom{n}{q}\sum_{k=1}^{q}\binom{n}{k}\binom{n-k}{q-k}\sum_{p=0}^{n-k}\binom{n-k}{p}(-1)^{n-k-p}\left(\binom{n}{q}-\binom{p}{q}\right)^{-2}.$$

We obtain nH_n when we evaluate this for $q = 1$ which is a good check but not exactly surprising since we have already seen this work at the other link (the reader is invited to attempt this computation using the above formula as a starting point, which is easier than what follows). We now try for a closed form for $q = 2$ and get

$$2n^2(n-1)^2\sum_{p=0}^{n-1}\binom{n-1}{p}(-1)^{n-1-p}\frac{1}{(n(n-1)-p(p-1))^2}$$

$$+n^2(n-1)^2\sum_{p=0}^{n-2}\binom{n-2}{p}(-1)^{n-2-p}\frac{1}{(n(n-1)-p(p-1))^2}.$$

We use residues as before with the function

$$g(z) = \frac{(n-1)!}{(n(n-1)-z(z-1))^2}\prod_{p=0}^{n-1}\frac{1}{z-p} = \frac{(n-1)!}{(z-n)^2(z-(1-n))^2}\prod_{p=0}^{n-1}\frac{1}{z-p}.$$

Note that

$$\left(\prod_{p=0}^{n-1} \frac{1}{z-p}\right)' = -\prod_{p=0}^{n-1} \frac{1}{z-p} \sum_{p=0}^{n-1} \frac{1}{z-p}$$

We get for the residue at n

$$(n-1)! \left(-\frac{1}{(2n-1)^2} \frac{1}{n!} H_n - \frac{2}{(2n-1)^3} \frac{1}{n!}\right)$$

The residue at $1-n$ yields

$$(n-1)! \left(\frac{1}{(1-2n)^2} \frac{(-1)^n (n-2)!}{(2n-2)!} (H_{2n-2} - H_{n-2}) - \frac{2}{(1-2n)^3} \frac{(-1)^n (n-2)!}{(2n-2)!}\right)$$

For the second sum we get for the residue at n

$$(n-2)! \left(-\frac{1}{(2n-1)^2} \frac{1}{n!} (H_n - 1) - \frac{2}{(2n-1)^3} \frac{1}{n!}\right)$$

and for the one at $1-n$

$$(n-2)! \left(\frac{1}{(1-2n)^2} \frac{(-1)^{n-1} (n-2)!}{(2n-3)!} (H_{2n-3} - H_{n-2}) - \frac{2}{(1-2n)^3} \frac{(-1)^{n-1} (n-2)!}{(2n-3)!}\right)$$

Collecting everything we find (observe that the terms on H_{2n-2} and on H_{n-2} and the third term from the residues at $1-n$ cancel the same way as in the computation of the probability that we saw earlier)

$$-\frac{1}{(2n-1)^2} H_n \left(\frac{2}{n} + \frac{1}{n(n-1)}\right) + \frac{1}{(2n-1)^2} \frac{1}{n(n-1)}$$
$$-\frac{2}{(2n-1)^3} \left(\frac{2}{n} + \frac{1}{n(n-1)}\right)$$
$$+(n-1)! \frac{2}{(1-2n)^2} \frac{(-1)^n (n-2)!}{(2n-2)!} \frac{1}{2n-2}$$

This is

$$-\frac{1}{(2n-1)n(n-1)} H_n - \frac{1}{(2n-1)^2} \frac{1}{n(n-1)}$$
$$+(n-2)! \frac{1}{(1-2n)^2} \frac{(-1)^n (n-2)!}{(2n-2)!}$$

Flip the sign and multiply by $n^2(n-1)^2$ to obtain the formula

$$\frac{n(n-1)}{2n-1} H_n + \frac{n(n-1)}{(2n-1)^2} + (-1)^{n-1} \frac{n^2(n-1)^2}{(2n-1)^2} \frac{(n-2)! \times (n-2)!}{(2n-2)!}$$

We conclude with the closed form (an **exact** result) for the case of packets containing two coupons which is

$$\frac{n(n-1)}{2n-1}H_n + \frac{n(n-1)}{(2n-1)^2} + (-1)^{n-1}\frac{n^2}{(2n-1)^2}\binom{2n-2}{n-1}^{-1}.$$

This attractive formula obviously motivates further research, possibly into the case $q = 3$, which looks difficult, perhaps requiring a computer algebra system during the simplifications. Observe that the dominant asymptotic is $1/2 \times nH_n$ which means $q = 2$ is about twice as fast as a single coupon so the effect of there being no packets with duplicate coupons is negligible.

This was math.stackexchange.com problem 2147576.

9 Expected time until the first k coupons where $k \le n$ have been collected

In the spirit of this document here is a generating function approach to the question of when the first k coupons where $k \le n$ have been seen. Using the notation from the following MSE link we get from first principles for the probability of m draws that

$$P[T=m] = \frac{1}{n^m} \times \binom{k}{k-1} \times \sum_{q=0}^{n-k} \binom{n-k}{q} \left\{ {m-1 \atop q+k-1} \right\} (q+k-1)!.$$

What happens here is that we choose the $k-1$ of the k values that go into the prefix, which also determines the value that will complete the set with the last draw. We then choose a set of q values not from the k initial ones and partition the first $m-1$ draws or slots into $q+k-1$ sets, one for each value.

We verify that this is a probability distribution, getting

$$\sum_{m\ge 1} P[T=m]$$

$$= \sum_{m\ge 1} \frac{1}{n^m} \times \binom{k}{k-1} \times \sum_{q=0}^{n-k} \binom{n-k}{q} (m-1)![z^{m-1}](\exp(z)-1)^{q+k-1}$$

$$= k\sum_{m\ge 1} \frac{1}{n^m} \times (m-1)![z^{m-1}] \sum_{q=0}^{n-k} \binom{n-k}{q} (\exp(z)-1)^{q+k-1}$$

$$= k\sum_{m\ge 1} \frac{1}{n^m} \times (m-1)![z^{m-1}](\exp(z)-1)^{k-1}\exp(z(n-k))$$

$$= k!\sum_{m\ge 1} \frac{1}{n^m} \sum_{q=0}^{m-1} \binom{m-1}{q} \left\{ {q \atop k-1} \right\} (n-k)^{m-1-q}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \sum_{m\ge q+1} \binom{m-1}{q} \frac{1}{n^m} (n-k)^{m-1-q}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \frac{1}{n^{q+1}} \sum_{m\ge 0} \binom{m+q}{q} \frac{1}{n^m} (n-k)^{m}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \frac{1}{n^{q+1}} \frac{1}{(1-(n-k)/n)^{q+1}} = k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \frac{1}{k^{q+1}}.$$

Recall the OGF of the Stirling numbers of the second kind which says that

$$\left\{ {n \atop k} \right\} = [z^n] \prod_{p=1}^{k} \frac{z}{1-pz}.$$

We obtain

$$k! \sum_{q\ge 0} \frac{1}{k^{q+1}} [z^q] \prod_{p=1}^{k-1} \frac{z}{1-pz} = (k-1)! \prod_{p=1}^{k-1} \frac{1/k}{1-p/k}$$

$$= (k-1)! \prod_{p=1}^{k-1} \frac{1}{k-p} = 1$$

and the sanity check goes through. For the expectation of when the first k have been seen we recycle the above, inserting a factor of m, starting from

$$k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \sum_{m\ge q+1} \binom{m-1}{q} \frac{m}{n^m} (n-k)^{m-1-q}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} \sum_{m\ge q+1} \binom{m}{q+1} \frac{q+1}{m} \frac{m}{n^m} (n-k)^{m-1-q}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} (q+1) \sum_{m\ge q+1} \binom{m}{q+1} \frac{1}{n^m} (n-k)^{m-1-q}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} (q+1) \frac{1}{n^{q+1}} \sum_{m\ge 0} \binom{m+q+1}{q+1} \frac{1}{n^m} (n-k)^{m}$$

$$= k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} (q+1) \frac{1}{n^{q+1}} \frac{1}{(1-(n-k)/n)^{q+2}}$$

$$= n \times k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} (q+1) \frac{1}{k^{q+2}} = \frac{n}{k^2} \times k! \sum_{q\ge 0} \left\{ {q \atop k-1} \right\} (q+1) \frac{1}{k^{q}}.$$

Activating the OGF produces

$$\frac{n}{k^2} \times k! \sum_{q\ge 0} \frac{1}{k^q} [z^q] \left(\prod_{p=0}^{k-1} \frac{z}{1-pz} \right)'$$

$$= \frac{n}{k^2} \times k! \sum_{q\ge 0} \frac{1}{k^q} [z^q] \prod_{p=0}^{k-1} \frac{z}{1-pz} \sum_{p=0}^{k-1} \frac{1}{z(1-pz)}$$

$$= \frac{n}{k^2} \times k! \prod_{p=0}^{k-1} \frac{1/k}{1-p/k} \sum_{p=0}^{k-1} \frac{1}{1/k(1-p/k)}$$

$$= n \times k! \prod_{p=0}^{k-1} \frac{1}{k-p} \sum_{p=0}^{k-1} \frac{1}{k-p} = n \times k! \times \frac{1}{k!} \times H_k.$$

This yields the answer

$$nH_k.$$

What we have here are in fact two *annihilated coefficient extractors* (ACE) more of which may be found at this MSE link.

This was math.stackexchange.com problem 2181969.

10 No replacement with j instances of each type of coupon

We use the notation from the following MSE link with m for the number of trials and n for the number of different types of coupons. We treat the special case where there are j instances of each type and we are sampling without replacement.

We ask about the probability of obtaining the distribution

$$\prod_{q=1}^{n} C_q^{\alpha_q}$$

where α_q says we have that many instances of type C_q. We obtain

$$\frac{(nj - \sum_{q=1}^{n} \alpha_q)!}{(nj)!} \prod_{q=1}^{n} \frac{j!}{(j-\alpha_q)!}.$$

Therefore the sequences according to probability of length m of n types of coupons without replacement and a maximum of j coupons of each type are given by

$$m![z^m] \left(\sum_{k=0}^{j} \frac{j!}{(j-k)!} \frac{z^k}{k!} \right)^n = m![z^m](1+z)^{nj} = \binom{nj}{m} \times m!.$$

Here we are partitioning the draws into n sets, one for each type, with $z^k/k!$ representing the size of the set and $j!/(j-k)!$ the weight according to probability.

Note also that $(nj)^{\underline{m}}$ gives the denominators of the probabilities while $j^{\underline{k}}$ gives the numerators corresponding to a set of size k.

We then obtain from first principles the formula

$$P[T=m] = \frac{1}{m!}\binom{nj}{m}^{-1} \times n \times j \times (m-1)![z^{m-1}]\left(\sum_{k=1}^{j} \frac{j!}{(j-k)!}\frac{z^k}{k!}\right)^{n-1}$$

$$= nj \times \frac{1}{m}\binom{nj}{m}^{-1}[z^{m-1}]\left(-1+(1+z)^j\right)^{n-1}.$$

This becomes

$$nj \times \frac{1}{m}\binom{nj}{m}^{-1}[z^{m-1}]\sum_{q=0}^{n-1}\binom{n-1}{q}(-1)^{n-1-q}(1+z)^{qj}$$

$$= \binom{nj-1}{m-1}^{-1}\sum_{q=0}^{n-1}\binom{n-1}{q}(-1)^{n-1-q}\binom{qj}{m-1}.$$

Observe that

$$\binom{qj}{m-1}\binom{nj-1}{m-1}^{-1} = \frac{(qj)! \times (nj-1-(m-1))!}{(qj-(m-1))! \times (nj-1)!}$$

$$= \binom{nj-1}{qj}^{-1}\binom{nj-1-(m-1)}{qj-(m-1)}.$$

We record for the probabilities the formula

$$P[T=m] = \sum_{q=0}^{n-1}\binom{n-1}{q}(-1)^{n-1-q}\binom{nj-1}{qj}^{-1}\binom{nj-1-(m-1)}{nj-1-qj}.$$

Start by verifying that this is a probability distribution. We obtain for the sum in m

$$\sum_{m=n}^{(n-1)j+1}\binom{nj-1-(m-1)}{nj-1-qj}$$

$$= [z^{nj-1-qj}]\sum_{m=n}^{(n-1)j+1}(1+z)^{nj-1-(m-1)}$$

$$= [z^{nj-1-qj}]\sum_{m=j-1}^{n(j-1)}(1+z)^m = [z^{nj-qj}]((1+z)^{n(j-1)+1}-(1+z)^{j-1}).$$

We have $nj - qj \geq j$ so only the first term contributes and we obtain

$$\sum_m P[T=m] = \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} \binom{n(j-1)+1}{nj-qj}$$

$$= \frac{n(j-1)+1}{j} \sum_{q=0}^{n-1} \binom{n-1}{q} \frac{(-1)^{n-1-q}}{n-q} \binom{nj-1}{nj-qj-1}^{-1} \binom{n(j-1)}{nj-qj-1}$$

We get for the rightmost pair of binomial coefficients

$$\frac{(n(j-1))! \times (qj)!}{(nj-1)! \times (qj+1-n)!} = \binom{nj-1}{n-1}^{-1} \binom{qj}{n-1}$$

which yields for the sum

$$\frac{n(j-1)+1}{j} \binom{nj-1}{n-1}^{-1} \sum_{q=0}^{n-1} \binom{n-1}{q} \frac{(-1)^{n-1-q}}{n-q} \binom{qj}{n-1}$$

$$= \frac{n(j-1)+1}{nj} \binom{nj-1}{n-1}^{-1} \sum_{q=0}^{n-1} \binom{n}{q} (-1)^{n-1-q} \binom{qj}{n-1}$$

$$= \frac{n(j-1)+1}{nj} \binom{nj-1}{n-1}^{-1} \binom{nj}{n-1}$$

$$+ \frac{n(j-1)+1}{nj} \binom{nj-1}{n-1}^{-1} \sum_{q=0}^{n} \binom{n}{q} (-1)^{n-1-q} \binom{qj}{n-1}$$

$$= \frac{n(j-1)+1}{nj} \frac{nj}{nj+1-n}$$

$$+ \frac{n(j-1)+1}{nj} \binom{nj-1}{n-1}^{-1} [z^{n-1}] \sum_{q=0}^{n} \binom{n}{q} (-1)^{n-1-q} (1+z)^{qj}$$

$$= 1 - \frac{n(j-1)+1}{nj} \binom{nj-1}{n-1}^{-1} [z^{n-1}](1-(1+z)^j)^n$$

Now observe that $[z^{n-1}](1-(1+z)^j)^n = 0$ hence everything simplifies to

$$1$$

and we have a probability distribution.

Continuing with the **expectation** we have the following closed form:

$$\mathrm{E}[T] = \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} \sum_{m=n}^{(n-1)j+1} m \binom{nj-1-(m-1)}{nj-1-qj}.$$

By means of plotting strategy let us examine some of these. Here are the first few for eight types of coupons starting at $j = 1$:

$$8, \frac{76627}{6435}, \frac{76801}{5434}, \frac{7473667}{480675}, \frac{1318429}{79794}, \ldots$$

and here is the initial segment for ten types of coupons:

$$10, \frac{707825}{46189}, \frac{7008811}{380380}, \frac{266299459}{13042315}, \frac{182251913}{8360638}, \frac{748880445829}{32831263465}, \ldots$$

Careful inspection of these values reveals that we cannot hope for additional simplification when $j \geq 2$ because if it were possible it would have appeared in these sample values. We do see however that the case $j = 1$ should be possible, the value being n (we always finish after n draws if there is only one instance of each coupon).

We now do this calculation, which is trivial, but nontheless a useful sanity check, starting with

$$\sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{n-1}{q}^{-1} \sum_{m=n}^{n} m \binom{n-1-(m-1)}{n-1-q}$$

$$= \sum_{q=0}^{n-1} (-1)^{n-1-q} \times n \binom{n-1-(n-1)}{n-1-q}$$

$$= (-1)^{n-1-(n-1)} \times n \times \binom{0}{n-1-(n-1)} = n.$$

It certainly seems like a worthwhile challenge to prove that the closed form for $\mathrm{E}[T]$ is nH_n in the limit, which is confirmed by the numerical evidence.

We did verify the formula for the expectation in software, as follows. It really is quite remarkable that the output from this program is in **excellent agreement** with the closed form on all values that were tested.

This was math.stackexchange.com problem 2172876.

Simplification

The above computation can be simplified. Using a slightly different notation we have N coupons partitioned int C clusters of size c_j and we start by computing the probability that after M draws we have seen all types of coupons. Using the method from above we see that it is given by

$$\frac{1}{M!} \binom{N}{M}^{-1} \times M! [z^M] \prod_{j=1}^{C} \sum_{k=1}^{c_j} \frac{c_j!}{(c_j - k)!} \frac{z^k}{k!} = \binom{N}{M}^{-1} [z^M] \prod_{j=1}^{C} \sum_{k=1}^{c_j} \binom{c_j}{k} z^k.$$

This is

$$\binom{N}{M}^{-1} [z^M] \prod_{j=1}^{C} (-1 + (1+z)^{c_j}).$$

For the special case of all clusters having the same size j we get

$$\binom{N}{M}^{-1} [z^M] (-1 + (1+z)^j)^C = \binom{N}{M}^{-1} [z^M] \sum_{q=0}^{C} \binom{C}{q} (-1)^{C-q} (1+z)^{qj}.$$

This is

$$\binom{N}{M}^{-1} \sum_{q=0}^{C} \binom{C}{q} (-1)^{C-q} \binom{qj}{M}.$$

We can use this to compute the expected number of draws until a representative from every cluster has been seen. Note that the complementary probability counts draws where at least one type of cluster is missing, i.e. the number of draws until having seen all is more than M. Hence we get for the expectation

$$\mathrm{E}[T] = N - j + 1 - \sum_{M=0}^{N-j} \binom{N}{M}^{-1} \sum_{q=0}^{C} \binom{C}{q} (-1)^{C-q} \binom{qj}{M}.$$

As a sanity check when $j = 1$ the expectation should be C. We obtain

$$C - \sum_{M=0}^{C-1} \binom{C}{M}^{-1} \sum_{q=M}^{C} \binom{C}{q} (-1)^{C-q} \binom{q}{M}.$$

Now we have

$$\binom{C}{q}\binom{q}{M} = \frac{C!}{(C-q)! \times M! \times (q-M)!} = \binom{C}{M}\binom{C-M}{C-q}.$$

Substituting we find

$$C - \sum_{M=0}^{C-1} \binom{C}{M}^{-1} \binom{C}{M} \sum_{q=M}^{C} \binom{C-M}{C-q} (-1)^{C-q}$$

$$= C - \sum_{M=0}^{C-1} \sum_{q=0}^{C-M} \binom{C-M}{C-M-q} (-1)^{C-M-q} = C - \sum_{M=0}^{C-1} \sum_{q=0}^{C-M} \binom{C-M}{q} (-1)^q$$

$$= C - \sum_{M=0}^{C-1} 0 = C,$$

as claimed. Here we have used that $C - 1 \geq M$ or $C \geq M + 1 > M$.
This was math.stackexchange.com problem 3119266.

10.1 No replacement with j instances of each type of coupon, a fixed number of draws and the number of types seen

By means of explaining the construction we may study an example where we have a fixed number m of draws and we are interested in the probability that q different types are seen, which gives the marked generating function

$$\left(1 + u\sum_{k=1}^{j} \frac{j!}{(j-k)!}\frac{z^k}{k!}\right)^n.$$

We thus have for the probability of q different types

$$\frac{1}{m!}{nj \choose m}^{-1} \times m![z^m][u^q]\left(1 + u\sum_{k=1}^{j} \frac{j!}{(j-k)!}\frac{z^k}{k!}\right)^n$$

$$= \frac{1}{m!}{nj \choose m}^{-1} \times m![z^m]{n \choose q}\left(\sum_{k=1}^{j} \frac{j!}{(j-k)!}\frac{z^k}{k!}\right)^q$$

$$= {nj \choose m}^{-1}[z^m]{n \choose q}\left(-1 + (1+z)^j\right)^q$$

$$= {nj \choose m}^{-1}[z^m]{n \choose q}\sum_{p=0}^{q}{q \choose p}(-1)^{q-p}(1+z)^{jp}.$$

We conclude that the desired probability is given by

$${nj \choose m}^{-1}{n \choose q}\sum_{p=0}^{q}{q \choose p}(-1)^{q-p}{jp \choose m}.$$

Observing that

$${nj \choose m}^{-1}{jp \choose m} = \frac{(jp)! \times (nj-m)!}{(jp-m)! \times (nj)!}$$

we get the alternate form

$${n \choose q}\sum_{p=0}^{q}{q \choose p}(-1)^{q-p}{nj \choose pj}^{-1}{nj-m \choose pj-m}.$$

E.g. for 15 draws from 10 types of coupons with 3 instances of each we obtain the PGF

$$\frac{7\,u^5}{4308820} + \frac{945\,u^6}{861764} + \frac{16191\,u^7}{430882}$$
$$+ \frac{112023\,u^8}{430882} + \frac{416745\,u^9}{861764} + \frac{938223\,u^{10}}{4308820},$$

a result that is not accessible by enumeration, which was nonetheless implemented as a sanity check in the Maple code posted in the on-line version, where it was found to match the two closed forms on the values that were examined.

This was math.stackexchange.com problem 2683788.

10.2 No replacement with n instances of each type of coupon, a fixed number of draws and the number of types not seen

The following is the complementary porblem to the previous section but it includes the expectation of the number of different types seen. Here we have G types of coupons, n of each and draw a sample of s coupons. We have from first principles as before that the PGF in u with the coefficient on $[u^q]$ representing the probability of q different colors / coupons not being seen in a sample of size s is given by

$$\frac{1}{s!}\binom{nG}{s}^{-1} s![z^s]\left(u+\sum_{k=1}^{n}\frac{n!}{(n-k)!}\frac{z^k}{k!}\right)^G.$$

This simplifies to

$$\binom{nG}{s}^{-1}[z^s]\left(u+\sum_{k=1}^{n}\binom{n}{k}z^k\right)^G = \binom{nG}{s}^{-1}[z^s](u-1+(1+z)^n)^G.$$

As a sanity check we indeed have on setting $u = 1$

$$\binom{nG}{s}^{-1}[z^s](1+z)^{nG} = 1.$$

For example, with four colors and four instances each we get for six draws the distribution

$$\binom{16}{6}^{-1}[z^6](u-1+(1+z)^4)^4 = \frac{3\,u^2}{143}+\frac{60\,u}{143}+\frac{80}{143}.$$

where e.g. the last term gives the probability that none of the colors are missing. We cannot have three colors missing because that leaves only one color to cover all six draws, we have only four instances, however. With this PGF we can answer the question about *the probability that q colors are missing in a draw of s items*, which is

$$\binom{nG}{s}^{-1}[z^s][u^q](u-1+(1+z)^n)^G = \binom{nG}{s}^{-1}[z^s]\binom{G}{q}(-1+(1+z)^n)^{G-q}$$

$$= \binom{nG}{s}^{-1}[z^s]\binom{G}{q}\sum_{p=0}^{G-q}\binom{G-q}{p}(-1)^{G-q-p}(1+z)^{np}.$$

This yields for the probability

$$\binom{nG}{s}^{-1} \binom{G}{q} \sum_{p=0}^{G-q} \binom{G-q}{p} (-1)^{G-q-p} \binom{np}{s}$$

which is inclusion-exclusion.

Returning to the main question we thus have for the expectation of coupons that did not occur

$$\binom{nG}{s}^{-1} \frac{\partial}{\partial u} [z^s](u-1+(1+z)^n)^G \Big|_{u=1}$$

$$= \binom{nG}{s}^{-1} [z^s]\; G(u-1+(1+z)^n)^{G-1}\big|_{u=1} = \binom{nG}{s}^{-1} [z^s] G(1+z)^{n(G-1)}.$$

We get for the number of coupons that did occur

$$G - G\binom{nG}{s}^{-1} \binom{nG-n}{s}.$$

E.g. when we draw one coupon we obtain

$$G - G\frac{1}{nG}(nG-n) = G - G + G\frac{n}{nG} = 1$$

as expected. Also note that we obtain the value G when $s > nG - n$ (second binomial coeffcient is zero). This is because the maximum coverage with $G-1$ colors is $nG - n$ and with the next sample we must use the last missing color.

The expectation may also be computed by linearity of expectation.

This was math.stackexchange.com problem.

11 No replacement with j instances of each type until all of one type is seen

We solve the problem where we have j instances of each of n types of coupons and draw without replacement until we have seen all j coupons of some type. Using the notation from the following MSE link we introduce the marked generating function

$$\left(\sum_{k=0}^{j-2} \frac{j!}{(j-k)!} \frac{z^k}{k!} + jwz^{j-1} \right)^n.$$

The coefficient on $[z^m]$ here represents distributions of sequences of m draws from the n types according to probability, where the ones that occur $j-1$ times have been marked. Each of the latter may be augmented to a complete set of some color where the weight is one because $j-1$ coupons have already been drawn. As we only need the count we differentiate with respect to w and set $w = 1$, getting

$$n \times \left(\sum_{k=0}^{j-1} \frac{j!}{(j-k)!} \frac{z^k}{k!}\right)^{n-1} \times jz^{j-1}.$$

With the method from the linked post we thus obtain for the probability

$$P[T = m] = \frac{1}{m!} \binom{nj}{m}^{-1} (m-1)! [z^{m-1}] njz^{j-1} \left(\sum_{k=0}^{j-1} \frac{j!}{(j-k)!} \frac{z^k}{k!}\right)^{n-1}$$

$$= \frac{1}{m!} \binom{nj}{m}^{-1} \times n \times j \times (m-1)! [z^{m-1}] z^{j-1} (-z^j + (1+z)^j)^{n-1}.$$

Extracting the coefficient we find

$$\binom{nj-1}{m-1}^{-1} [z^{m-j}] \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} z^{j(n-1-q)} (1+z)^{qj}$$

$$= \binom{nj-1}{m-1}^{-1} \sum_{q=0}^{n-1} [z^{m-j(n-q)}] \binom{n-1}{q} (-1)^{n-1-q} (1+z)^{qj}$$

$$= \binom{nj-1}{m-1}^{-1} \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{qj}{m-j(n-q)}$$

$$= \binom{nj-1}{m-1}^{-1} \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{qj}{nj-m}.$$

Observe that

$$\binom{qj}{nj-m} \binom{nj-1}{m-1}^{-1} = \frac{(qj)!(m-1)!}{(nj-1)!(m-(n-q)j)!}$$

$$= \binom{nj-1}{qj}^{-1} \binom{m-1}{m-(n-q)j}.$$

We record for the probabilities the formula

$$P[T = m] = \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} \binom{m-1}{(n-q)j-1}.$$

We now **verify that this is a probability distribution.** This requires the value of

$$\sum_{m=j}^{n(j-1)+1} \binom{m-1}{(n-q)j-1} = \sum_{m=j-1}^{n(j-1)} \binom{m}{(n-q)j-1}$$

$$= [z^{(n-q)j-1}] \sum_{m=j-1}^{n(j-1)} (1+z)^m = [z^{(n-q)j}]((1+z)^{n(j-1)+1} - (1+z)^{j-1}).$$

With $0 \le q \le n-1$ the second term does not contribute and we may continue with

$$\sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} \binom{n(j-1)+1}{(n-q)j}$$

$$= \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} \binom{nj+1-n}{qj+1-n}$$

$$= \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \frac{qj}{nj-qj} \binom{nj-1}{qj-1}^{-1} \binom{nj+1-n}{qj+1-n}$$

$$= \sum_{q=1}^{n-1} \binom{n-1}{q-1} (-1)^{n-1-q} \binom{nj-1}{qj-1}^{-1} \binom{nj+1-n}{qj+1-n}.$$

Observe once more that

$$\binom{nj-1}{qj-1}^{-1} \binom{nj+1-n}{qj+1-n} = \frac{(nj+1-n)! \times (qj-1)!}{(nj-1)! \times (qj+1-n)!}$$

$$= \binom{nj-1}{n-2}^{-1} \binom{qj-1}{n-2}.$$

We thus find for the sum of the probabilities

$$\binom{nj-1}{n-2}^{-1} \sum_{q=1}^{n-1} \binom{n-1}{q-1} (-1)^{n-1-q} \binom{qj-1}{n-2}$$

$$= \binom{nj-1}{n-2}^{-1} \sum_{q=0}^{n-2} \binom{n-1}{q} (-1)^{n-q} \binom{qj+j-1}{n-2}$$

$$= 1 + \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-q} \binom{qj+j-1}{n-2}.$$

The sum vanishes, as in

$$\sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-q} [z^{n-2}] (1+z)^{qj+j-1}$$

$$= [z^{n-2}](1+z)^{j-1} \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-q} (1+z)^{qj}$$

$$= [z^{n-2}](1+z)^{j-1}(1-(1+z)^j)^{n-1},$$

but $(1-(1+z)^j)^{n-1} = (-1)^{n-1} j^{n-1} z^{n-1} + \cdots$ and there is no contribution. This confirms it being a probability distribution.

Continuing with the expectation we require the value of

$$\sum_{m=j}^{n(j-1)+1} m \binom{m-1}{(n-q)j-1} = (n-q)j \sum_{m=j}^{n(j-1)+1} \binom{m}{(n-q)j}$$

$$= (n-q)j[z^{(n-q)j}] \sum_{m=j}^{n(j-1)+1} (1+z)^m$$

$$= (n-q)j[z^{(n-q)j+1}]((1+z)^{n(j-1)+2} - (1+z)^j) = (n-q)j \binom{n(j-1)+2}{(n-q)j+1}.$$

The second term did not contribute since we have $(n-q)j+1 > j$. We thus have for the expectation

$$j \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} (n-q) \binom{n(j-1)+2}{(n-q)j+1}$$

$$= j \sum_{q=0}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj}^{-1} (n-q) \binom{nj+2-n}{qj+1-n}$$

$$= j \sum_{q=1}^{n-1} \binom{n-1}{q} (-1)^{n-1-q} \frac{qj}{nj-qj} \binom{nj-1}{qj-1}^{-1} (n-q)$$

$$\times \frac{nj+2-n}{nj-qj+1} \binom{nj+1-n}{qj+1-n}$$

$$= j(nj+2-n) \sum_{q=1}^{n-1} q \binom{n-1}{q} (-1)^{n-1-q} \binom{nj-1}{qj-1}^{-1} \frac{1}{nj-qj+1} \binom{nj+1-n}{qj+1-n}.$$

Re-using the earlier factorization we get

$$j(nj+2-n) \binom{nj-1}{n-2}^{-1} \sum_{q=1}^{n-1} q \binom{n-1}{q} (-1)^{n-1-q} \frac{1}{nj-qj+1} \binom{qj-1}{n-2}$$

$$= j(nj+2-n)(n-1) \binom{nj-1}{n-2}^{-1} \sum_{q=1}^{n-1} \binom{n-2}{q-1} (-1)^{n-1-q} \frac{1}{nj-qj+1} \binom{qj-1}{n-2}$$

$$= j^2 n(nj+1)\binom{nj+1}{n-1}^{-1}$$

$$\times \sum_{q=0}^{n-2} \binom{n-2}{q} (-1)^{n-q} \frac{1}{nj-qj-j+1} \binom{qj+j-1}{n-2}.$$

Working with the sum term we have

$$\binom{n-2}{q} (-1)^{n-q} \frac{1}{nj-qj-j+1} \binom{qj+j-1}{n-2}$$

$$= \mathrm{Res}_{z=q} \frac{1}{nj-j+1-zj} \prod_{p=0}^{n-3} (zj+j-1-p) \prod_{p=0}^{n-2} \frac{1}{z-p}.$$

Now since $\lim_{R\to\infty} 2\pi R \times R^{n-2}/R/R^{n-1} = 0$ and residues sum to zero we may evaluate this by taking the negative of the residue at $z = n-1+1/j$. This is the computation:

$$-\mathrm{Res}_{z=n-1+1/j} \frac{1}{nj-j+1-zj} \prod_{p=0}^{n-3} (zj+j-1-p) \prod_{p=0}^{n-2} \frac{1}{z-p}$$

$$= \frac{1}{j} \mathrm{Res}_{z=n-1+1/j} \frac{1}{z-(n-1+1/j)} \prod_{p=0}^{n-3} (zj+j-1-p) \prod_{p=0}^{n-2} \frac{1}{z-p}$$

$$= \frac{1}{j} \prod_{p=0}^{n-3} (nj-p) \prod_{p=0}^{n-2} \frac{1}{n-1+1/j-p}$$

$$= \frac{1}{j} \times (n-2)! \times \binom{nj}{n-2} \times j^{n-1} \prod_{p=0}^{n-2} \frac{1}{nj-j+1-pj}.$$

With

$$\binom{nj}{n-2}\binom{nj+1}{n-1}^{-1} = \frac{n-1}{nj+1}$$

we finally have the closed form

$$E[T] = n! \times j^n \times \prod_{p=0}^{n-2} \frac{1}{nj-j+1-pj}.$$

To see what the asymptotics are we use the alternate form

$$E[T] = n \times j \times \frac{\Gamma(n)\Gamma(1+1/j)}{\Gamma(n+1/j)}.$$

Keeping j fixed and letting n go to infinity yields the asymptotic

$$n \times j \times \Gamma(1+1/j) \times n^{-1/j} = n^{1-1/j} \times j \times \Gamma(1+1/j).$$

There is an enumeration routine that may be compared to the closed forms both of which were implemented in the Maple code at the on-line version of this answer.

With the calculation that was presented here we want to make sure we have the correct interpretation of the problem from the start. The following basic program will do this by computing the expectation through simulation. Consult for the details of the scenario under investigation. The output is in fine agreement with the data i.e. the closed form from above.

This was math.stackexchange.com problem 2401573.

12 No replacement with j instances of each type until two of one type is seen

Introductory remark. This answer was originally written to treat the case of seeing the first repeated value when drawing without replacement from a deck of cards as proposed at this MSE link. It answers the corresponding case of n pairs of socks as well, however, consult the end of the document for this.

We solve the problem where we have j instances of each of n types of coupons and draw without replacement until we have seen 2 coupons of some type. For a deck of cards we have 13 types of coupons and 4 instances of each type. Using the notation from the following MSE link I and MSE link II we introduce the marked generating function

$$(1 + jwz)^n .$$

The coefficient on $[z^m]$ here represents distributions of sequences of m draws from the n types according to probability, where the ones that occur one time have been marked. Each of the latter may be augmented to a pair of some color where the weight is $j - 1$ because one coupon has already been drawn. As we only need the count we differentiate with respect to w and set $w = 1$, getting

$$n \times (1 + jz)^{n-1} \times jz.$$

With the method from the linked posts we thus obtain for the probability

$$\begin{aligned} P[T = m] &= \frac{1}{m!} \binom{nj}{m}^{-1} (m-1)! \times (j-1) \times [z^{m-1}] njz(1+jz)^{n-1} \\ &= \frac{nj}{m} \binom{nj}{m}^{-1} \times (j-1) \times [z^{m-2}](1+jz)^{n-1} \\ &= \frac{nj}{m} \binom{nj}{m}^{-1} \times (j-1) \times \binom{n-1}{m-2} j^{m-2} \\ &= (j-1) \times \binom{nj-1}{m-1}^{-1} \binom{n-1}{m-2} j^{m-2}. \end{aligned}$$

Next we **verify that this is a probability distribution.** The process may halt after two steps at the earliest and $n+1$ at the latest and we get

$$\sum_{m=2}^{n+1} P[T=m] = (j-1)\sum_{m=2}^{n+1} \binom{nj-1}{m-1}^{-1} \binom{n-1}{m-2} j^{m-2}$$
$$= (j-1)\sum_{m=2}^{n+1} \binom{nj-1}{m-1}^{-1} \frac{m-1}{n} \binom{n}{m-1} j^{m-2}$$
$$= \frac{j-1}{n}\sum_{m=2}^{n+1} \binom{nj-1}{m-1}^{-1} \binom{n}{m-1} (m-1) j^{m-2}.$$

We have

$$\binom{nj-1}{m-1}^{-1} \binom{n}{m-1} = \frac{n! \times (nj-m)!}{(nj-1)! \times (n-(m-1))!}$$
$$= \binom{nj-1}{n}^{-1} \binom{nj-m}{n-(m-1)}.$$

Here we have used the fact that for the scenario to make sense we must have $j \geq 2$. Continuing we find

$$\frac{j-1}{n}\binom{nj-1}{n}^{-1} \sum_{m=2}^{n+1} \binom{nj-m}{n-(m-1)} (m-1) j^{m-2}$$

The sum term yields

$$\sum_{m=1}^{n} \binom{nj-1-m}{n-m} m j^{m-1} = \sum_{m\geq 1} [w^{n-m}](1+w)^{nj-1-m} m j^{m-1}$$
$$= [w^n](1+w)^{nj-1} \sum_{m\geq 1} w^m (1+w)^{-m} m j^{m-1}$$
$$= [w^n](1+w)^{nj-1} \frac{w}{(1+w)} \sum_{m\geq 1} w^{m-1} (1+w)^{-(m-1)} m j^{m-1}$$
$$= [w^n](1+w)^{nj-1} \frac{w}{(1+w)} \frac{1}{(1-wj/(1+w))^2}$$
$$= [w^{n-1}](1+w)^{nj} \frac{1}{(1+w-wj)^2} = [w^{n-1}](1+w)^{nj} \frac{1}{(1-(j-1)w)^2}.$$

Extracting coefficients we find

$$\sum_{q=0}^{n-1} \binom{nj}{n-1-q} (q+1)(j-1)^q = \sum_{q=0}^{n-1} \binom{nj}{nj-n+q+1} (q+1)(j-1)^q$$

$$= nj \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q} (j-1)^q - n(j-1) \sum_{q=0}^{n-1} \binom{nj}{nj-n+q+1} (j-1)^q$$

$$= n \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q} (j-1)^{q+1} + n \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q} (j-1)^q$$

$$-n \sum_{q=0}^{n-1} \binom{nj}{nj-n+q+1} (j-1)^{q+1}$$

$$= n \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q} (j-1)^{q+1} + n \sum_{q=-1}^{n-2} \binom{nj-1}{nj-n+q+1} (j-1)^{q+1}$$

$$-n \sum_{q=0}^{n-1} \binom{nj}{nj-n+q+1} (j-1)^{q+1}$$

$$= n \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q} (j-1)^{q+1} + n \sum_{q=0}^{n-1} \binom{nj-1}{nj-n+q+1} (j-1)^{q+1}$$

$$+n\binom{nj-1}{nj-n} - n \sum_{q=0}^{n-1} \binom{nj}{nj-n+q+1} (j-1)^{q+1} = n\binom{nj-1}{nj-n}.$$

Collecting everything we obtain

$$\frac{j-1}{n} \binom{nj-1}{n}^{-1} \times n \times \binom{nj-1}{n-1}$$

$$= \frac{j-1}{n} \binom{nj-1}{n}^{-1} \times n \times \binom{nj-1}{n} \frac{n}{nj-n} = 1$$

and we have confirmed that we have a probability distribution.

The next step is to **compute the expectation.** Recapitulating the earlier computation we find that

$$E[T] = \sum_{m=2}^{n+1} mP[T=m] = \frac{j-1}{n} \binom{nj-1}{n}^{-1} \sum_{m=1}^{n} \binom{nj-1-m}{n-m} (m+1)mj^{m-1}$$

or

$$E[T] = \frac{2(j-1)}{n} \binom{nj-1}{n}^{-1} [w^{n-1}](1+w)^{nj+1} \frac{1}{(1-(j-1)w)^3}.$$

Extracting coefficients we obtain the closed form

$$E[T] = \frac{(j-1)}{n}\binom{nj-1}{n}^{-1}\sum_{q=0}^{n-1}\binom{nj+1}{n-1-q}(q+2)(q+1)(j-1)^q.$$

Observe that for a deck of cards we get

$$E[T] = \frac{226087256246}{39688347475} \approx 5.696565129.$$

Furthermore this simplifies when $j = 2$ (pairs of socks). Instantiating j to 2 will produce

$$\frac{2}{n}\binom{2n-1}{n}^{-1}[w^{n-1}](1+w)^{2n+1}\frac{1}{(1-w)^3}.$$

The coefficient is

$$\mathrm{Res}_{w=0}\frac{1}{w^n}(1+w)^{2n+1}\frac{1}{(1-w)^3}.$$

Note that the residue at infinity is given by

$$-\mathrm{Res}_{w=0}\frac{1}{w^2}w^n\frac{(1+w)^{2n+1}}{w^{2n+1}}\frac{1}{(1-1/w)^3} = -\mathrm{Res}_{w=0}\frac{1}{w^2}\frac{(1+w)^{2n+1}}{w^{n+1}}\frac{w^3}{(w-1)^3}$$
$$= \mathrm{Res}_{w=0}\frac{(1+w)^{2n+1}}{w^n}\frac{1}{(1-w)^3}.$$

Hence the value is minus half the residue at $w = 1$. We find with $(1-w)^3 = -(w-1)^3$

$$\frac{1}{2}\times\frac{1}{2}\frac{1}{w^n}(1+w)^{2n+1}\left(\frac{n(n+1)}{w^2} - \frac{2n(2n+1)}{w(1+w)} + \frac{(2n+1)(2n)}{(1+w)^2}\right)\Bigg|_{w=1}$$
$$= 2^{2n-1}\left(n^2+n-2n^2-n+n^2+\frac{1}{2}n\right) = \frac{1}{4}n4^n.$$

Now observe that

$$\binom{2n-1}{n}^{-1} = \binom{2n}{n}^{-1}\times 2n\times\frac{1}{n} = 2\binom{2n}{n}^{-1} \sim 2\times\frac{\sqrt{\pi n}}{4^n}$$

We thus have the closed form for $j = 2$

$$E[T] = \binom{2n-1}{n}^{-1}\frac{1}{2}4^n = \binom{2n}{n}^{-1}4^n.$$

and we get the nice asymptotic

$$E[T] \sim \sqrt{\pi n}.$$

There is also a very basic C program which confirmed the closed form of the expectations for all combinations of n and j that were examined. For example with $j = 5$ we get the expectations

$$2, \frac{23}{9}, \frac{272}{91}, \frac{3253}{969}, \frac{6522}{1771}, \frac{94477}{23751}, \frac{714436}{168113}, \frac{69263329}{15380937}, \ldots$$

with values

$$2, 2.555555556, 2.989010989, 3.357069143, 3.682665161,$$

$$3.977811461, 4.249736784, 4.503193076, \ldots$$

Running the program on 10^8 trials will then match these values to about five digits decimal precision.

This was math.stackexchange.com problem 2453824.

13 No replacement with $\binom{n}{j}$ instances of each type of coupon

This problem is a type of coupon collector without replacement where there are $\binom{n}{j}$ tickets of type j and we ask about the expectation of the sum of the ticket values after m tickets have been drawn. Using the methodology from the following two MSE links we find that the EGF by multiplicity of a set of coupons of type j is given by

$$\sum_{k=0}^{\binom{n}{j}} \binom{n}{j}^{\underline{k}} \frac{z^k}{k!} = (1+z)^{\binom{n}{j}}.$$

Distributing all n types of coupons we get

$$\prod_{j=0}^{n} (1+z)^{\binom{n}{j}} = (1+z)^{2^n}$$

for a total count according to multiplicity of

$$m![z^m](1+z)^{2^n} = m! \times \binom{2^n}{m}.$$

Marking the contribution of a ticket of type j with u^j we obtain the mixed generating function

$$G(z, u) = \prod_{j=0}^{n} (1 + u^j z)^{\binom{n}{j}}.$$

Differentiate and evaluate at $u = 1$ to obtain

$$\left.\frac{\partial}{\partial u} G(z,u)\right|_{u=1}$$

$$= \left.\prod_{j=0}^{n}(1+u^j z)^{\binom{n}{j}} \sum_{j=0}^{n}(1+u^j z)^{-\binom{n}{j}} \binom{n}{j} (1+u^j z)^{\binom{n}{j}-1} j u^{j-1} z\right|_{u=1}$$

$$= (1+z)^{2^n} \sum_{j=1}^{n} \binom{n}{j} \frac{jz}{1+z} = z(1+z)^{2^n-1} \sum_{j=1}^{n} j \binom{n}{j}$$

$$= z(1+z)^{2^n-1} \sum_{j=1}^{n} n \binom{n-1}{j-1} = n2^{n-1} z(1+z)^{2^n-1}.$$

Extracting coefficients we thus obtain for the expectation of the sum

$$\mathrm{E}[S] = \binom{2^n}{m}^{-1} n2^{n-1} \binom{2^n-1}{m-1} = n2^{n-1}\frac{m}{2^n} = \frac{1}{2}nm.$$

Continuing with the variance we evidently require the second factorial moment. Differentiating twice we get three components, the first is

$$\left.\prod_{j=0}^{n}(1+u^j z)^{\binom{n}{j}} \sum_{j=0}^{n}(1+u^j z)^{-\binom{n}{j}} \binom{n}{j} (1+u^j z)^{\binom{n}{j}-1} j(j-1) u^{j-2} z\right|_{u=1}$$

$$= z(1+z)^{2^n-1} \sum_{j=2}^{n} j(j-1) \binom{n}{j}$$

$$= z(1+z)^{2^n-1} \sum_{j=2}^{n} n(n-1) \binom{n-2}{j-2} = n(n-1)2^{n-2} z(1+z)^{2^n-1}.$$

The second is

$$\left.\prod_{j=0}^{n}(1+u^j z)^{\binom{n}{j}} \sum_{j=0}^{n}(1+u^j z)^{-\binom{n}{j}} \binom{n}{j} \left(\binom{n}{j}-1\right) (1+u^j z)^{\binom{n}{j}-2} j^2 u^{2j-2} z^2\right|_{u=1}$$

$$= z^2(1+z)^{2^n-2} \sum_{j=1}^{n} j^2 \binom{n}{j} \left(\binom{n}{j}-1\right).$$

The third is

$$2\prod_{j=0}^{n}(1+u^j z)^{\binom{n}{j}} \sum_{j=0}^{n}(1+u^j z)^{-\binom{n}{j}} \binom{n}{j} (1+u^j z)^{\binom{n}{j}-1} j u^{j-1} z$$

$$\times \sum_{k=j+1}^{n} (1+u^k z)^{-\binom{n}{k}} \binom{n}{k} (1+u^k z)^{\binom{n}{k}-1} k u^{k-1} z \Bigg|_{u=1}$$

$$= 2z^2(1+z)^{2^n-2} \sum_{j=0}^{n} \binom{n}{j} j \sum_{k=j+1}^{n} \binom{n}{k} k.$$

The coefficients on these last two may be joined and we get

$$-\sum_{j=1}^{n} j^2 \binom{n}{j} + \left(\sum_{j=1}^{n} j \binom{n}{j} \right)^2$$

$$= -\sum_{j=1}^{n} j(j-1)\binom{n}{j} - \sum_{j=1}^{n} j \binom{n}{j} + \left(n \sum_{j=1}^{n} \binom{n-1}{j-1} \right)^2$$

$$= -n(n-1)\sum_{j=2}^{n} \binom{n-2}{j-2} - n \sum_{j=1}^{n} \binom{n-1}{j-1} + n^2 2^{2n-2}$$

$$= -n(n-1)2^{n-2} - n2^{n-1} + n^2 2^{2n-2} = n^2 2^{2n-2} - n(n+1)2^{n-2}.$$

Extracting coefficients we get for the second factorial moment

$$\frac{1}{4} n(n-1)m + (n^2 2^{2n-2} - n(n+1)2^{n-2}) \frac{m(m-1)}{2^n(2^n-1)}$$

or alternatively

$$\mathrm{E}[S(S-1)] = \frac{1}{4} n(n-1)m + \frac{1}{4}\frac{m(m-1)}{2^n-1}(n^2 2^n - n(n+1)).$$

Finally recall that

$$\mathrm{Var}[S] = \mathrm{E}[S(S-1)] + \mathrm{E}[S] - \mathrm{E}[S]^2$$

so the answer to the problem posed by the OP is

$$\mathrm{E}[S] = \frac{1}{2} nm$$

and

$$\mathrm{Var}[S] = \frac{1}{4} n(n+1)m + \frac{1}{4}\frac{m(m-1)}{2^n-1}(n^2 2^n - n(n+1)) - \frac{1}{4} n^2 m^2.$$

As a sanity check when $m = 2^n$ and all coupons have been drawn we have deterministically that

$$\mathrm{E}[S] = \sum_{j=0}^{n} j \binom{n}{j} = n \sum_{j=1}^{n} \binom{n-1}{j-1} = n2^{n-1} = \frac{1}{2} nm$$

and the check goes through.

With this problem requiring careful algebra I also coded a simulation of the coupon collector that is featured in the on-line version which was in **excellent** agreement on all values that were tested (outputs first and second factorial moment). Some optimizations are still possible which is left as an exercise to the reader.

This was math.stackexchange.com problem 2362256.

References

[Ego84] G.P. Egorychev. *Integral Representation and the Computation of Combinatorial Sums.* American Mathematical Society, 1984.

[Wik17] Wikipedia. Coupon collector's problem — wikipedia, the free encyclopedia, 2017. [Online; accessed 18-February-2017].