Probability in the Engineering and Informational Sciences, **27**, 2013, 141–146.
doi:10.1017/S026996481200037X

A BALANCE SCALE PROBLEM

Mark Brown
Department of Mathematics
The City College, CUNY, NY
E-mail: mbrown@ccny.cuny.edu

Suppose we have n objects of different weights. We randomly sample pairs of objects, and for each sampled pair use a balance scale to determine which of the two objects is heavier. It is assumed that the sequence of sampled pairs is iid, each selection uniformly distributed on the set of $n(n-1)/2$ pairs. We continue sampling until the first time that we can definitively identify the heaviest of the n objects. The problem of interest is to compute the expected number of selected pairs.

1. INTRODUCTION AND SUMMARY

Balance scale problems have generated considerable interest. Many are presented and discussed online and can be accessed through Google. The following problem was brought to my attention by Professor Robert Blau of Baruch College, to whom I am most grateful.

A balance scale can determine which of the 2 objects is heavier. It does not provide weights or differences in weight. Suppose we have n objects and we randomly sample pairs, (i, j). The sequence of sampled pairs are assumed i.i.d., each uniformly distributed among the $\binom{n}{2}$ pairs. At each stage we may use the balance scale to determine which of the two chosen objects is heavier, if we do not already know from previous comparisons. It is assumed that all objects differ in weight and that the balance scale does a perfect job of comparisons. The problem is to compute ET_n, where T_n is the random number of selections made until we determine which of the objects is the heaviest.

Our main result is that,

$$\frac{ET_n}{\binom{n}{2}}$$

is an inceasing sequence with limit c, where

$$1.254768 < c < 1.255575. \tag{1}$$

The value of the constant c can be pinned down further by our methodology.

The above problem can be viewed as a coupon collecting problem. It can also be interpreted in terms of exponential order statistics. If $X_1 \ldots X_n$ are i.i.d. exponentially distributed with mean 1, and $Y_1 < Y_2 \cdots < Y_n$ are the corresponding order statistics, then we will derive,

$$\frac{ET_{n+1}}{\binom{n+1}{2}} = E\left[\max_{1\le j\le n}(Y_j - Y_{j-1})\right] = E\left[\max_{1\le j\le n}\frac{X_j}{j}\right], \quad \text{and}$$

$$c = \lim_{n\to\infty} E\left[\max_{1\le j\le n}(\frac{X_j}{j})\right].$$

Maximums of independent exponential random variables arise in reliability (Barlow and Proschan [1], p. 83), in coupon collecting models (Ross [3], p. 314), and in relaxation times for a class of time reversible Markov chains (Brown [2], p. 378).

The expected maximum can be expressed as a sum of $2^n - 1$ terms, but this is hardly computationally useful for large n. Section (4.6) of Pekoz and Ross [4], points out and addresses this problem, suggesting clever techniques for obtaining bounds. The approach followed here is more accurate for this particular case, but involves more computation than the Pekoz–Ross approach.

2. DERIVATION OF RESULTS

Suppose that item i_1 is the heaviest, i_2 the second heaviest, and so on until item i_n, the lightest. Define,

$$C_r = \{(i_1, i_r), (i_2, i_r) \ldots, (i_{r-1}, i_r)\}, \quad r = 2, \ldots, n.$$

We know which of the items are heaviest as soon as we have sampled at least one member from each of $C_2, \ldots, C_n$. This is because any comparison in C_r tells us that item i_r is not the heaviest. We do not know which item is heaviest prior to sampling at least one member from each of $C_2 \ldots C_n$. For example if we have no comparison in C_r, then even with all the remaining comparison pairs we cannot conclude that i_1 is heavier than i_r.

Now, $P(C_r) = (r-1)/\binom{n}{2}, r = 2, \ldots n$.

The problem thus reduces to one of multinomial trials with, $p_r = P(C_r)$, $r = 2, \ldots, n$, in which we sample until we obtain one observation from each category. This is equivalent to the coupon collecting problem with coupon probabilities, $p_r, r = 2, \ldots, n$.

Consider $n-1$ independent Poisson processes with rates $p_r, r = 2, \ldots n$. The superimposed Poisson process has rate 1. Each time an event occurs in the superimposed process, the probability that it arose in process r, equals p_r, independently of all other events. Thus, the discrete time coupon collecting problem is replaced by a continuous time process for which the mean time between coupons equals one. The discrete and continuous time mean waiting times (until at least one of each type) are identical.

Let $\{\varepsilon_i, i \ge 1\}$ denote i.i.d. exponential random variables with mean 1. Then,

$$\begin{aligned} ET_n &= E\,\max\left(\frac{\varepsilon_j}{p_j}, j = 2, \ldots, n\right) = E\max\left(\binom{n}{2}\frac{\varepsilon_\ell}{\ell}, \ell = 1, \ldots, n-1\right) \\ &= \binom{n}{2} E\,\max(\frac{\varepsilon_\ell}{\ell}, \ell = 1, \ldots, n-1). \end{aligned} \quad \textbf{(2)}$$

https://doi.org/10.1017/S026996481200037X Published online by Cambridge University Press

Define $Z_n = \max_{1\leq j\leq n}(\frac{\varepsilon_j}{j}), \mu_n = EZ_n$. Note that $ET_{n+1} = \binom{n+1}{2}\mu_n$. Next,

$$\begin{aligned}\mu_n - \mu_{n-1} &= E\left[I_{\frac{\varepsilon_n}{n}>Z_{n-1}}\left(\frac{\varepsilon_n}{n} - Z_{n-1}\right)\right]\\ &= Pr(\frac{\varepsilon_n}{n} > Z_{n-1})E\left(\frac{\varepsilon_n}{n} - Z_{n-1}|\frac{\varepsilon_n}{n} > Z_{n-1}\right)\\ &= \frac{1}{n}Pr(\frac{\varepsilon_n}{n} > Z_{n-1}) \quad \text{(lack of memory property).}\end{aligned} \tag{3}$$

Defining, $\gamma_j = Pr(\varepsilon_j/j > Z_{j-1})$ with $\gamma_1 = 1$, then from (3),

$$\mu_n = \sum_{j=1}^{n}\frac{\gamma_j}{j}. \tag{4}$$

Define, $A_j = \{(\varepsilon_j/j) > (\varepsilon_n/n)\}$, $j = 1, \ldots, n$.

It follows that, $\gamma_n = 1 - Pr(U_1^n A_j)$, By the inclusion–exclusion formula (Ross [3], p. 584),

$$\frac{\gamma_n}{n} = \sum_{\alpha\epsilon B_{n-1}}(-1)^{|\alpha|}\frac{1}{n+s_\alpha} \tag{5}$$

where B_{n-1} is the collection of the 2^{n-1} subsets of $\{1, \ldots, n-1\}$, α is a subset of $\{1, \ldots, n-1\}$, with cardinality $|\alpha|$, and $S_\alpha = \sum_{j\epsilon\alpha} j$. As,

$$Pr(Z_{n-1} \leq t) = \prod_{1}^{n-1}(1 - e^{-jt}) = \sum_{B_{n-1}}(-1)^{|\alpha|}e^{-S_\alpha t}$$

it follows that,

$$\frac{L_{n-1}^{(s)}}{s} \stackrel{\text{def}}{=} \frac{Ee^{-sZ_{n-1}}}{s} = \sum_{B_{n-1}}(-1)^{|\alpha|}\frac{1}{S_\alpha + s}. \tag{6}$$

From (5) and (6),

$$\frac{\gamma_n}{n} = \frac{L_{n-1}^{(n)}}{n}. \tag{7}$$

Next,

$$L_n(s) = Ee^{-sZ_n} = \sum_{B_n}(-1)^{|\alpha|}\frac{1}{S_\alpha + s} = \sum_{B_{n-1}}(-1)^{|\alpha|}\frac{1}{S_\alpha + s} - \sum_{B_{n-1}}(-1)^{|\alpha|}\frac{1}{S_\alpha + s + n}.$$

Thus,

$$L_n(s) = \frac{L_{n-1}^{(s)}}{s} - \frac{L_{n-1}^{(s+n)}}{s+n}. \tag{8}$$

From (7) and (8),

$$\frac{\gamma_n}{n} = \frac{L_{n-1}^{(n)}}{n} = \frac{L_{n-2}^{(n)}}{n} - \frac{L_{n-2}^{(2n-1)}}{2n-1}. \tag{9}$$

Now, $L_0(s)/s = \frac{1}{s}$, and $a(s) \stackrel{\text{def}}{=} Ł_1(s)/s = 1/s(s+1)$. Then, $E(Z_2 - Z_1) = \gamma_2/2 = L_0(2)/2 - L_0(3)/3 = a(2) = 1/6$

$$E(Z_3 - Z_2) = \frac{\gamma_3}{3} = \frac{L_1(3)}{3} - \frac{L_1(5)}{5} = a(3) - a(5) = \frac{1}{20}. \tag{10}$$

https://doi.org/10.1017/S026996481200037X Published online by Cambridge University Press

TABLE 1. $EZ_j, j = 1, 2 \ldots 12$

j	EZ_j
1	1
2	$7\|6 = 1.166667$
3	$73\|60 = 1.216667$
4	$89\|72 = 1.236111$
5	1.244902
6	1.249295
7	1.251653
8	1.252989
9	1.253780
10	1.254265
11	1.254571
12	1.254769

We continue in this manner. To go from γ_n/n to $(\gamma_{n+1})/(n+1)$, we replace a term $a(k)$ appearing in γ_n/n by $a(k+1) - a(k+n)$. For example from (10) we replace $a(3)$ by $a(4) - a(7)$ and $a(5)$ by $a(6) - a(9)$. This yields,

$$\frac{\gamma_4}{4} = a(4) - a(6) - a(7) + a(9) = \frac{7}{360} \tag{11}$$

Applying this procedure to (11) we obtain,

$$\frac{\gamma_5}{5} = a_5 - a_7 - a_8 - a_9 + a_{10} + a_{11} + a_{12} - a_{14} = \frac{4}{455}.$$

For n even, the terms a_k and $a(\binom{n+2}{2} - k - 2)$ have the same coefficients. For n odd, their coefficients are of opposite sign with the same absolute values. The next term is,

$$\begin{aligned}\frac{\gamma_6}{6} &= a(6) - a(8) - a(9) - a(10) + a(12) + 2(a(13)) \\ &\quad + a(14) - a(16) - a(17) - a(18) + a(20) \\ &= 0.0043923182. \end{aligned} \tag{12}$$

Continuing in this manner we compute EZ_j for $j = 1, 2, \ldots, 12$. The values in the table 1 are rounded off to 6 decimal places.

Thus, for example,

$$ET_{13} = \binom{13}{2} EZ_{12} \approx 97.871954.$$

Remark. The order statistics interpretion for, $ET_{n+1}/\binom{n+1}{2}$, follows because $\{Y_j - Y_{j-1}, j = 1, \ldots, n\}$ are independent exponential random variables with parameters, $\{n - j + 1, j = 1, \ldots, n\}$, and thus $\max(Y_j - Y_{j-1})$ is distributed as the maximum of n i.i.d. exponentials with parameter 1. It follows that for Yule process with parameter 1 (Ross [3], p. 377), the constant c can be interpreted as the mean of, $T = \max_n T_n$, where T_n is the waiting time to go from state n to state $n+1$. I tried to compute c from the distribtion of T, but had no success.

3. UPPER BOUND

For $n \geq 14$ we know that $ET_n > \binom{n}{2} EZ_{12}$ but we do not yet have an upper bound. To obtain an upper bound observed that for $r < n$,

$$\gamma(r, n) \stackrel{\text{def}}{=} Pr\left(\frac{\varepsilon_n}{n} > Z_r\right) \geq Pr\left(\frac{\varepsilon_n}{n} > Z_{n-1}\right) = \gamma_n.$$

Consequently for $n \geq m+1 > r$

$$E(Z_n - Z_m) = \sum_{j=m+1}^{n} \frac{\gamma_j}{j} \leq \sum_{j=m+1}^{n} \frac{\gamma(r,j)}{j}.$$

For the choice $r = 6$ we compute, for $j > 6$,

$$\begin{aligned}\frac{\gamma(6,j)}{j} = \frac{L_6(j)}{j} &= \frac{1}{j} - \frac{1}{j+1} - \frac{1}{j+2} + \frac{1}{j+5} + \frac{2}{j+7} - \frac{1}{j+9} - \frac{1}{j+10} \\ &\quad - \frac{1}{j+11} - \frac{1}{j+12} + \frac{2}{j+14} + \frac{1}{j+16} - \frac{1}{j+19} - \frac{1}{j+20} + \frac{1}{j+21}. \end{aligned} \tag{13}$$

For $n \geq 13$,

$$E(Z_n - Z_{12}) = \sum_{13}^{n} \frac{\gamma_j}{j} \leq \sum_{13}^{n} \frac{\gamma(6,j)}{j}.$$

Define, $H_m = \sum_1^m \frac{1}{j}$, then $\sum_{j=13}^{n} [1/(j+k)] = H_{n+k} - H_{12+k}$. Applying the summation to each term in (13), and collecting like terms we obtain,

$$\sum_{13}^{n} \frac{\gamma(6,j)}{j} = f(12) - f(n)$$

where

$$\begin{aligned} f(k) = \frac{1}{k+1} &- \left(\frac{1}{k+3} + \frac{1}{k+4} + \frac{1}{k+5}\right) + 2\left(\frac{1}{k+8} + \frac{1}{k+9}\right) + \frac{1}{k+10} \\ &- \frac{1}{k+12} - 2\left(\frac{1}{k+13} + \frac{1}{k+14}\right) + \left(\frac{1}{k+17} + \frac{1}{k+18} + \frac{1}{k+19}\right) - \frac{1}{k+21}, \end{aligned}$$

with f positive and decreasing to 0 as $k \to \infty$.

TABLE 2. Possible log convexity of d_a, and log concavity of e_n

n	d_n	e_n
1	0.166667	1.800000
2	0.300000	1.296296
3	0.388889	1.162593
4	0.452119	1.105076
5	0.499626	1.074455
6	0.536826	1.056004
7	0.566890	1.043935
8	0.591797	1.035564
9	0.612844	1.029299
10	0.630799	1.025332

$$d_n = \frac{\gamma_{n+1/n+1}}{\gamma_{n/n}},$$

$$e_n = \frac{d_{n+1}}{d_n}. \tag{14}$$

https://doi.org/10.1017/S026996481200037X Published online by Cambridge University Press

Consequently, for $n \geq 13$,

$$EZ_{12} < EZ_n < EZ_{12} + f(12) - f(n) < EZ_{12} + f(12).$$

As $f(12) = 0.0008069041$ and $EZ_{12} = 1.254769$ it follows that for $n \geq 13$ that,

$$1.254769 < EZ_n < 1.255576$$

and that $c = \lim EZ_n$ falls in this same range. Thus, result (1) has been obtained.

Additional Remarks

Based on the calculation of $\gamma_j/j, j = 1, \ldots, 12$, numerical calculations suggest the possibility that,

$$d_n \stackrel{\text{def}}{=} \frac{\gamma_{n+1}/(n+1)}{\gamma_n/n}$$

is increasing in n and thus that γ_n/n is a decreasing log convex sequence.

It also appears plausible that

$$e_n \stackrel{\text{def}}{=} \frac{d_{n+1}}{d_n}$$

is decreasing in n, and thus that d_n is an increasing log concave sequence.

I pose the problem of further investigating these possibilities. Perhaps the complete monotonicity property of Laplace transforms is relevant to the solution. Table 2 provides some numerical evidence.

References

1. Barlow, R.E. & Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models.* New York: Holt, Rinehart and Winston, Inc.
2. Brown, M. (1975). The first passage time distribution for a parallel exponential system with repair. In *Reliability and fault tree Analysis*, R.E. Barlow, et al. (eds) Philadelphia, PA: SIAM. pp. 365–396.
3. Ross, S.M. (2007). *Introduction to probability models*, 9th ed. Amsterdam: Elsevier.
4. Ross, S.M. & Pekoz, E. (2007). *A second course in probability.* Boston, MA: www.ProbabilityBookstore.com.

https://doi.org/10.1017/S026996481200037X Published online by Cambridge University Press