Competition between Discrete Random Variables, with Applications to Occupancy Problems

Julia Eaton
University of Washington, USA

Anant P. Godbole
East Tennessee State University, USA

D. Elizabeth "Betsy" Sinclair
Caltech/University of Chicago, USA

May 15, 2007

1 Introduction

A preliminary version of this abstract was submitted to IWAP, Piraeus, 2003. Consider n players whose "scores" are independent and identically distributed values $\{X_i\}_{i=1}^n$ from some discrete distribution F. We consider the case of general distributions F but pay special attention to the cases where (i) F is geometric with parameter $p \to 0$ and (ii) F is uniform on $\{1, 2, \ldots, N\}$; the latter case corresponds to the classical occupancy problem. The quantities of interest to us are

- the U-statistic W which counts the number of "ties" between pairs i, j (with $X_a = X_b = X_c = X_d$, for example, leading to a contribution of $\binom{4}{2} = 6$ to the value of W);
- the univariate statistic Y_r which counts the number of strict r-way ties between contestants, i.e., episodes of the form $X_i = x$ for some x iff $i \in A$, $|A| = r$; and

- the multivariate vector $Z_b = (Y_2, Y_3, \ldots, Y_b)$.

We provide Poisson approximations for the distributions of W, Y_r and Z_b under some general conditions. New results on the joint distribution of cell counts in the occupancy problem are derived as a corollary.

Consider the following game (see [?]): "Two players use a coin that lands heads with probability p to play a game that consists of a sequence of rounds. In each round, the first player tosses the coin until a head appears. Then the second player tosses the coin until a head appears. If the players have the same number of flips in a round, the round is declared a tie and another round is played. If not, the player with the larger number of flips wins the game. Rounds are played successively until one of the two players wins the game." Readers are asked to find the expected number of rounds; the expected value of the total number of flips; and the probability distribution of the difference between the number of flips made by players 1 and 2 in a given round. We briefly mention the solution for the first two of these questions: The probability of a two person tie is clearly

$$\sum_{x=1}^{\infty}(1-p)^{2x-2}p^2 = \frac{p}{2-p},$$

so that $\mathbb{E}(R)$, the expected number of rounds is given by

$$\mathbb{E}(R) = \sum_{x=1}^{\infty} x(\mathbb{P}(\text{tie}))^{x-1}(1-\mathbb{P}(\text{tie})) = \sum_{x=1}^{\infty} x\left(\frac{p}{2-p}\right)^{x-1}\left(1-\frac{p}{2-p}\right) = \frac{2-p}{2-2p},$$

so that Wald's lemma yields for $\mathbb{E}(F)$, the expected total number of flips,

$$\mathbb{E}(F) = \mathbb{E}(F/R)\mathbb{E}(R) = \frac{2-p}{2-2p}\mathbb{E}(F/R) = \frac{2(2-p)}{p(2-2p)},$$

since the expected number $\mathbb{E}(F/R)$ of flips per round is clearly $2/p$. Computations for a three-person game, not mentioned in [?], are similar, but we need to lay down some rules as follows: Three players each flip a p-coin until heads is flipped. The player with the highest number of flips wins unless there are ties between one or more players, in which case we repeat the process. That is, the value of each of the three geometric variables in question must be unique. We next compute the probability of a two- or three-way tie; the expected number of rounds; and the expected number of flips for $n = 3$ –

to convince the reader that the situation rapidly becomes quite complicated as n increases.

With three players (A,B,C), there are 3!=6 ways to have a strict inequality and seven ways to tie, since there is one way for a three way tie (which we loosely write as "A = B = C") to occur; $\binom{3}{1} = 3$ ways for A > B = C to occur; and $\binom{3}{2} = 3$ ways for A = B > C to occur. Note that

$$\begin{aligned}\mathbb{P}(A = B = C) &= p^3 + p^3(1-p)^3 + p^3(1-p)^6 \ldots \\ &= \sum_{x=1}^{\infty} p^3(1-p)^{3x-3} \\ &= \frac{p^2}{3 - 3p + p^2},\end{aligned}$$

while the table below

Case	A	B	C
1	TH, TTH, ...	H	H
2	TTH, TTTH, ...	TH	TH
3	TTTH, TTTTH, ...	TTH	TTH
4	TTTTH, TTTTTH, ...	TTTH	TTTH
⋮	⋮	⋮	⋮

reveals that

$$\mathbb{P}(A > B = C) = p^3 \sum_{m=1}^{\infty} (1-p)^m \sum_{i=0}^{m-1} (1-p)^{2i} = \frac{p(1-p)}{3 - 3p + p^2}.$$

Finally, we observe from the table

Case	A	B	C
1	TH	TH	H
2	TTH	TTH	H, TH
3	TTTH	TTTH	H, TH, TTH
4	TTTTH	TTTTH	H, TH, TTH, TTTH
⋮	⋮	⋮	⋮

that

$$\mathbb{P}(A = B > C) = p^3 \sum_{m=1}^{\infty} (1-p)^{2m} \sum_{i=0}^{m-1} (1-p)^{i} = \frac{p(1-p)^2}{(2-p)(3 - 3p + p^2)},$$

which leads to

$$\begin{aligned}\mathbb{P}(\text{tie}) &= \mathbb{P}(A=B=C)+3\mathbb{P}(A>B=C)+3\mathbb{P}(A=B>C)\\ &= \frac{5p^3-13p^2+9p}{(2-p)(3-3p+p^2)},\end{aligned}$$

and hence as before to

$$\mathbb{E}(R)=\frac{1}{1-\frac{5p^3-13p^2+9p}{(2-p)(3-3p+p^2)}}$$

and

$$\mathbb{E}(F)=\mathbb{E}(F/R)\mathbb{E}(R)=\frac{3}{p}\left(\frac{1}{1-\frac{5p^3-13p^2+9p}{(2-p)(3-3p+p^2)}}\right).$$

Competitions of the kind discussed above are best formulated in the more general context of occupancy models as follows: n balls are independently thrown into an infinite array of boxes so that any ball hits the jth box with probability p_j. Let X_j be the number of balls in box j. Then, with $p_j=(1-p)^{j-1}p$, we have the game inspired by [?] ending iff $X_j\le 1\ \forall j$. Extremal versions of such questions have arisen in the literature before, often with surprising results. Motivated by a question, posed by Carl Pomerance and arising in an additive number theory context, Athreya and Fidkowski [?] proved that the probability π_n that the highest numbered non-empty box has exactly one ball in it converges to a constant (which is shown to be one) iff $\lim_{n\to\infty} p_n/\sum_{j=n}^{\infty} p_j=0$. This is a condition that is not satisfied by, e.g., the sequence $p_n=1/2^n$ for which, quite interestingly, the limit superior and the limit inferior of the sequence π_n differ in the *fourth* decimal place. The authors of [?] found out after their work was accepted, however, that their results had been obtained a few years earlier by Eisenberg et. al [?], [?], [?] and also by Bruss and O'Cinneide [?]. The comprehensive paper of Móri [?] is most relevant too: Here it is proven that given a double sequence of integer valued random variables, i.i.d. within rows, and letting $\mu(n)$ denote the multiplicity of the maximal value in the nth row, the limiting distribution of $\mu(n)$ does not exist in the ordinary sense – but that the intriguing empirical type a.s. limit result

$$\lim_{t\to\infty}\frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}I(\mu(n)=m)=\frac{r^m}{m\log\left(\frac{1}{1-r}\right)}, m=1,2,\ldots$$

holds, where r is a parameter that depends on the distribution. The whole field appears to be extraordinarily rich with known facts and tantalizing possibilities.

Results of the kind described above sent us a message. We realized that our n person coin game was confounded with the fact that "cell counts" were unlikely to behave in an asymptotically smooth way if $p = p_n \not\to 0$. We thus made $p \to 0$ a blanket assumption when continuing with the analysis of the n-person game. In Theorem 1, we study the distribution of the number W of pairs of equalities in the n person game, with $W = 0$ corresponding to the end of a "round" in the sense of [**?**], and show that a good Poisson approximation obtains. Theorem 2 concerns itself with the distribution of the number Y_r of strict r-way ties (=the number of boxes with exactly r balls) and Theorem 3 with a multivariate generalization of Theorem 2. The approximating distribution is Poisson (Theorem 2) or a product of independent Poisson variates (Theorem 3). We note, moreover, that we were able to prove a result such as Theorem 3 *probably* due to the approach taken – which uses as a counter the event that r *specific* balls go into the same urn, rather than the conventional approach (e.g., [**?**], Section 6.2) of counting the number of urns with r balls. Two additional references relevant to this paper are [**?**] and [**?**]. The comprehensive body of work done by Arratia, Barbour and Tavaré in [**?**],[**?**], and [**?**] is focused on *decomposable logarithmic combinatorial structures*; our model, on the other hand, is decomposable but not logarithmic.

References

[1] R. Arratia, A. Barbour, and S. Tavaré (1999). "On Poisson-Dirichlet limits for random decomposable combinatorial structures," *Comb. Probab. Computing* **8**, 193–208.

[2] R. Arratia, A. Barbour, and S. Tavaré (2000). "Limits of logarithmic combinatorial structures," *Ann. Probab.* **28**, 1620–1644.

[3] R. Arratia, A. Barbour, and S. Tavaré (2001). *Logarithmic Combinatorial Structures.* Unpublished manuscript.

[4] R. Arratia, L. Goldstein, and L. Gordon (1989). "Two moments suffice for Poisson approximation: the Chen-Stein method," *Ann. Probab.* **17**, 9–25.

[5] J. Athreya and L. Fidkowski (2000). "Number Theory, Balls in Boxes, and the Asymptotic Uniqueness of Maximal Discrete Order Statistics," *Electronic Journal of Combinatorial Number Theory* **0**, Paper A3, available at `http://www.integers-ejcnt.org/vol0.html`.

[6] A. Barbour, L. Holst, and S. Janson (1992). *Poisson Approximation.* Oxford University Press.

[7] Y. Baryshnikov, B. Eisenberg, and G. Stengle (1995). "A necessary and sufficient condition for the existence of the limiting probability of a tie for first place," *Statist. Probab. Lett.* **23**, 203–209.

[8] F. Bruss and C. O'Cinneide (1990). "On the maximum and its uniqueness for geometric random samples," *J. Appl. Probab.* **22**, 598–610.

[9] B. Eisenberg, G. Stengle, and G. Strang (1993). "The asymptotic probability of a tie for first place," *Ann. Appl. Probab.* **3**, 731–745.

[10] B. Eisenberg and G. Stengle (1996). "Minimizing the probability of a tie for first place," *J. Math. Anal. Appl.* **198**, 458–472.

[11] V. Kolchin, B. Sevast'yanov, and V. Chistyakov (1978). *Random Allocations.* Winston, Washington, DC.

[12] T. Móri (2000). "On the multiplicity of the sample maximum and the longest head run," *Periodica Mathematica Hungarica* **41**, 195–212.

[13] B. Warner and B. Kline (1999). Problem Number 662, *College Mathematics Journal* **30**, 407.