Moments of Select Sets*

Simon Langowski† Mark Daniel Ward‡

Abstract

We analyze a selection procedure introduced by Krieger, Pollak, and Samuel–Cahn. It retains an item if it is among the top $100p$ percent, as compared to the items that have been accepted so far. Gaither and Ward analyzed the average behavior of the number of items selected. We present the asymptotic properties of the higher moments of the number of items retained by the selection procedure. We derive a general formula for the moments. To demonstrate the complexity of these moments, we present the exact first-order asymptotic growth of some of these moments, for various rational values of p.

1 Introduction

Krieger et al. introduced a natural selection algorithm for which an item is selected if it is in the top $100p$ percentile of items, compared to all previously selected items. Such select sets of items chosen by this selection algorithm could be applied to model quality control (acceptances and rejections) on an assembly line. These selection algorithms arise squarely in the context of the family of secretary problems, which have been studied since the 1960's. For a starting point to the vast literature on the secretary problem, we suggest [1].

The selection rule of Krieger et al. requires only that the rankings of all items are all equally likely. This assumption is standard in the study of the secretary problem; see, for instance, [1]. We emphasize that the standard of acceptance or rejection for an item is stochastic: The threshold for acceptance fundamentally depends on the previous items (again, an item is selected if it is in the top $100p$ percentile of items, *compared to all previously selected items*). Since the threshold for acceptance or rejection is developed in real time, this selection algorithm is quite realistic and could actually be utilized in practice, in situations that demand dynamic quality control.

Having the information about the moments can further inform a decision maker about the potential use of this algorithm, and (in particular) and the selection of the parameter p, namely, the percentile for the cutoff.

We sketch the analysis of the moments of the number of items retained by the selection algorithm. We follow the notation of Krieger et al. Among the first n items to be considered, we let L_n denote the number of items that are selected. To get the selection procedure started, we always accept the first item, so $L_1 = 1$. During the selection procedure, the $(n+1)$st item is retained if and only if its rank among the first $n+1$ items is among the best $100p$ percentile of L_n; this happens with probability $\lceil pL_n \rceil/(n+1)$. This discretization (introduced by the floor function) is a key source of the challenge in the analysis. Our key result is that the mth moment of the number of items L_n retained by the selection algorithm have (asymptotic) behavior of the form

$$\lim_{n\to\infty} E((L_n)^m)/n^{mp} = c_{m,p}.$$

In particular, for rational values of p, we can derive representations of $c_{m,p}$ that depend on transcendental numbers and special functions.

Krieger et al. [2] showed that

$$\lim_{n\to\infty} E(L_n)/n^p = c_p,$$

where c_p is a constant that depends only on p, but they stated that, "It seems impossible to determine c_p analytically, except for $p = 1$." This gap was the motivation for the second author's investigation of the first moments, in the earlier paper [3]. In this paper, we perform the analysis of all moments of L_n.

As a short recap of the previous paper [3]: The authors utilized the ordinary generating function of the expected value of L_n, then setup a differential equation to characterize this generating function, and utilized the behavior of this generating function near $z = 1$ (the closest singularity to the origin). As a result, [3] provided a first-order asymptotic characterization of $E(L_n)$. In the present paper, we generalize such a methodology, so that we are able to characterize the asymptotic growth of all moments of L_n.

*The authors' research are supported by NSF DMS-1246818. Ward's research is also supported by NSF Science & Technology Center for Science of Information Grant CCF-0939370.

†Department of Statistics, Purdue University, West Lafayette, IN 47907, USA

‡Department of Statistics, Purdue University, West Lafayette, IN 47907, USA

2 Main Results

Our main result characterizes the asymptotics of the moments of L_n, i.e., of the number of chosen items among the first n items in the selection algorithm.

THEOREM 2.1. *For any percentile* $100p$ *and moment* m, *we have*

$$\lim_{n\to\infty} E((L_n)^m)/n^{mp} = c_{m,p}$$

where $c_{m,p}$ *is a constant that depends only on* m *and* p, *namely,*

(2.1)

$$c_{m,p} = \frac{1 + \sum\limits_{k\geq 1} \left((k+1)^m - k^m - \frac{mpk^m}{\lceil pk\rceil}\right) \prod\limits_{j=1}^{k} \left(\frac{1}{1+\frac{mp}{\lceil pj\rceil}}\right)}{(mp+1)\Gamma(mp+1)}.$$

For rational values of p, these constants can be evaluated with the aid of a computer algebra system.

Intuitively, theorem 2.1 states that the mth moment of L_n asymptotically grows at a rate proportional to n^{mp}, and it also gives the precise leading constant for this first-order asymptotic term.

It should be noted that for $m = 1$, the constants c_p from Krieger et al. [2] and from Gaither and Ward [3] are exactly the constants $c_{1,p}$ in the present paper, since they are relevant for the first moments' asymptotics. Moreover, Theorem 1 of Gaither and Ward [3] is the special case $m = 1$ of our new (more general) Theorem 2.1.

The tables at the end of the paper illustrate exact values of $c_{m,p}$ for $m = 1, 2, 3$, and for selected values of p.

In general, we consider rational p of the form $p = r/s$. As a special case, we note that, when $p = 1/s$ or $p = 1 - 1/s$, the form of $c_{m,p}$ simplifies to

$$c_{m,1/s} = \frac{m^{s-1}}{m+s} s^{m-s+2}\Gamma(m/s)^{s-1},$$

or (respectively) to

$$c_{m,(s-1)/s} = \frac{m}{m(s-1)+s}\left(\frac{s}{s-1}\right)^m (s-1)^{m/s}\Gamma(m/s).$$

This simplification is due to the cycles in the remainder of the ceiling function. This can be seen by substituting $k = ls + b$ into equation (2.1). This substitution yields the following form

$$c_{m,p} = \frac{1 + \sum_{l\geq 0}(\prod_{\sigma=1}^{l} \mu_{m,r,s}(\sigma))(\sum_{b=1}^{s} \nu_{m,r,s}(l,b))}{(mp+1)\Gamma(mp+1)},$$

where we are using the notation

$$\mu_{m,r,s}(\sigma) := \prod_{j=1}^{s} \frac{1}{1+\frac{mp}{(\sigma-1)r+\lceil pj\rceil}},$$

and

$$\nu_{m,r,s}(l,b) := \left((ls+b+1)^m - (ls+b)^m - \frac{mp(ls+b)^m}{rl+\lceil pb\rceil}\right) \times \prod_{i=1}^{b} \frac{1}{1+\frac{mp}{rl+\lceil pi\rceil}}.$$

The representations $\mu_{m,r,s}$ and $\nu_{m,r,s}(l,b)$ generalize the analogous representations in Gaither and Ward [3]. See the proofs in Section 3 for more details.

One reason for the simplification in the case when $p = 1/s$ is that for each i and b in the summations and products seen above, we have $\lceil pi\rceil = \lceil pb\rceil = 1$. A similar simplifying argument in the case $p = 1 - 1/s$ holds for all i and b in the analogous derivations.

For other values of p, the derivations are much more complicated.

For the graphs in Figure 1, the coefficients on the moments were plotted for values of $1000 \leq j \leq 10000$ where $p = j/10000$.

From these graphs, we see that $c_{m,p}$ is (generally trending) in such a way that $c_{m,p}$ decreases as p increases, but $c_{m,p}$ is not continuous in p. Indeed, as a function of p (and for m fixed), we note that $c_{m,p}$ is discontinuous at each rational value of p.

3 Sketch of Proof of Theorem 2.1

3.1 Existence First, it is necessary to show $\lim_{n\to\infty} E(L_n^m)/n^{mp}$ exists and is finite. Since x^m is convex (for $m > 1$), then $(c_{1,p})^a = (E(L_n))^a < E(L_n^a)$, giving a lower bound for each moment by Jensen's inequality, as the first moment was found in [3]. To prove the upper bound, we note that the selection algorithm proceeds in steps, in a recursive manner. If the nth item is retained, then $L_n = L_{n-1} + 1$; otherwise, if the nth item is rejected, then $L_n = L_{n-1}$. This results in a recurrence for the mth moment:

$$(3.2)\qquad E(L_n^m \mid L_{n-1}) = L_{n-1}^m\Big(1 - \frac{\lceil pL_{n-1}\rceil}{n}\Big) + (L_{n-1}+1)^m \frac{\lceil pL_{n-1}\rceil}{n}.$$

Expanding the term $(L_{n-1}+1)^m$ yields

$$E(L_n^m \mid L_{n-1}) = L_{n-1}^m + \sum_{j=0}^{m-1} \binom{m}{j} L_{n-1}^j \frac{\lceil pL_{n-1}\rceil}{n},$$

Then we can bound the ceiling function, to obtain

(3.3)

$$E(L_n^m \mid L_{n-1}) \leq L_{n-1}^m + \sum_{j=0}^{m-1} \binom{m}{j} L_{n-1}^j \frac{pL_{n-1}+1}{n}.$$

By algebraic manipulation, we observe

$$\begin{aligned}
&\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j}\frac{pL_{n-1}+1}{n}\\
&=\frac{p}{n}\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j+1}+\frac{1}{n}\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j}\\
&=\frac{p}{n}\sum_{j=1}^{m}\binom{m}{j-1}L_{n-1}^{j}+\frac{1}{n}\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j}.
\end{aligned}$$

Now we isolate the L_{n-1}^{m} term, to obtain

$$\begin{aligned}
&\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j}\frac{pL_{n-1}+1}{n}\\
&=\frac{mp}{n}L_{n-1}^{m}+\frac{p}{n}\sum_{j=1}^{m-1}\binom{m}{j-1}L_{n-1}^{j}+\frac{1}{n}\sum_{j=0}^{m-1}\binom{m}{j}L_{n-1}^{j}\\
&=\frac{mp}{n}L_{n-1}^{m}+\frac{1}{n}\sum_{j=0}^{m-1}h_jL_{n-1}^{j}.
\end{aligned}$$

where the h_j's are constants that depend only on p but not on n, namely: $h_j := p\binom{m}{j-1}+\binom{m}{j}$ for $1 \le j \le m-1$, and $h_0 := 1$. Substituting into equation (3.3) yields

$$E(L_n^m \mid L_{n-1}) \le \Big(1+\frac{mp}{n}\Big)L_{n-1}^{m}+\frac{1}{n}\sum_{j=0}^{m-1}h_jL_{n-1}^{j}.$$

It follows, by taking the expected value with respect to L_{n-1}, that

$$E(L_n^m) \le \Big(1+\frac{mp}{n}\Big)E(L_{n-1}^{m})+\frac{1}{n}\sum_{j=0}^{m-1}h_jE(L_{n-1}^{j}).$$

To simplify the following derivations, we define

$$U_{n,m} := E(L_n^m)/n^{mp}.$$

Dividing the previous equation by n^{mp}, we obtain

$$U_{n,m} \le \Big(1+\frac{mp}{n}\Big)\frac{1}{n^{mp}}E(L_{n-1}^{m})+\frac{1}{n^{mp+1}}\sum_{j=0}^{m-1}h_jE(L_{n-1}^{j}).$$

which simplifies to

$$\begin{aligned}
U_{n,m} &\le \Big(1+\frac{mp}{n}\Big)\Big(1-\frac{1}{n}\Big)^{mp}U_{n-1,m}\\
&\quad+\frac{1}{n^{mp+1}}\sum_{j=0}^{m-1}h_jE(L_{n-1}^{j}).
\end{aligned}\tag{3.4}$$

We note that

$$\begin{aligned}
\frac{1}{n^{mp+1}}\sum_{j=0}^{m-1}h_jE(L_{n-1}^{j}) &= \sum_{j=0}^{m-1}\frac{(n-1)^{jp}}{n^{mp+1}}h_jU_{n-1,j}\\
&\le \sum_{j=0}^{m-1}\frac{n^{jp}}{n^{mp+1}}h_jU_{n-1,j}\\
&= \sum_{j=0}^{m-1}\frac{h_jU_{n-1,j}}{n^{(m-j)p+1}}.
\end{aligned}$$

Substituting into (3.4) yields

$$\begin{aligned}
U_{n,m} &\le \Big(1+\frac{mp}{n}\Big)\Big(1-\frac{1}{n}\Big)^{mp}U_{n-1,m}\\
&\quad+\sum_{j=0}^{m-1}\frac{h_jU_{n-1,j}}{n^{(m-j)p+1}}.
\end{aligned}\tag{3.5}$$

Equivalently, we have

$$\begin{aligned}
U_{n,m}-U_{n-1,m} &\le \Big(\Big(1+\frac{mp}{n}\Big)\Big(1-\frac{1}{n}\Big)^{mp}-1\Big)U_{n-1,m}\\
&\quad+\sum_{j=0}^{m-1}\frac{h_jU_{n-1,j}}{n^{(m-j)p+1}}.
\end{aligned}\tag{3.6}$$

As in Krieger et al. [2], note that the function

$$(1-x)^{mp}(1+mpx)-1$$

is zero at $x=0$, and the function has negative derivative for $0<x<1$. When $x=1/n$, this implies that $(1-1/n)^{mp}(1+mp/n)-1 \le 0$ for $n \ge 1$.

For the second part, we need to use strong induction, and create the inductive hypothesis that $\lim_{n\to\infty} E(L_n^k)/n^{kp}$ (i.e., $\lim_{n\to\infty} U_{n,k}$) is a constant for each $1 \le k < m$. For the base case, Krieger et al. [2], already showed that this holds for the first two moments of L_n, i.e., that $\lim_{n\to\infty} U_{n,k}$ is a constant for $m=1,2$. Now we proceed with the strong induction. We assume that $\lim_{n\to\infty} U_{n,k}$ is a constant for $1 \le k < m$, and we prove that $\lim_{n\to\infty} U_{n,m}$ is a constant. In (3.6), we have

$$\begin{aligned}
U_{n,m}-U_{n-1,m} &\le \Big(\Big(1+\frac{mp}{n}\Big)\Big(1-\frac{1}{n}\Big)^{mp}-1\Big)U_{n-1,m}\\
&\quad+\sum_{j=0}^{m-1}\frac{h_jU_{n-1,j}}{n^{(m-j)p+1}}.
\end{aligned}\tag{3.7}$$

Since, by the inductive hypothesis, $\lim_{n\to\infty} U_{n-1,j}$ exists, then there is a constant H_j such that $h_jU_{n-1,j} \le$

H_j for all n. This yields

$$U_{n,m} - U_{n-1,m} \le \left(\left(1 + \frac{mp}{n}\right)\left(1 - \frac{1}{n}\right)^{mp} - 1\right)U_{n-1,m} \tag{3.8}$$
$$+ \sum_{j=0}^{m-1} \frac{H_j}{n^{(m-j)p+1}}.$$

For convenience, we define $U_{0,m} := 0$. Then we observe that

$$U_{n,m} = \sum_{i=1}^{n}(U_{i,m} - U_{i-1,m}).$$

Now we apply (3.8) (using i instead of n) to each of the differences $U_{i,m} - U_{i-1,m}$. This gives us

$$U_{n,m} \le \sum_{i=1}^{n} \Big[\left(\left(1 + \frac{mp}{i}\right)\left(1 - \frac{1}{i}\right)^{mp} - 1\right)U_{i-1,m} + \sum_{j=0}^{m-1} \frac{H_j}{i^{(m-j)p+1}}\Big].$$

We already proved that the first part of the right hand side is negative. We also know that

$$\sum_{i=1}^{\infty} \frac{H_j}{i^{(m-j)p+1}}$$

exists for each j, since $(m-j)p + 1 > 1$. This gives an upper bound on $U_{n,m}$ (for fixed m) over all n.

Finally we use the sequence $U_{n,m}$ to show that the limit is not only bounded but exists by showing that the sequence will increase until it hits the upper bound.

A similar argument holds, using $\lceil pL_{n-1}/n\rceil \ge pL_{n-1}/n$, to establish a lower bound, and the lower bound is on the differences. This similar argument completes the strong induction.

3.2 Analysis For succinctness, we use $P_{n,k}$ to denote $P(L_n = k)$. The differences of the mth moments for L_{n+1} and L_n are

$$E(L_{n+1}^m) - E(L_n^m) = \sum_{k\ge1}(k^m P_{n+1,k} - k^m P_{n,k}), \tag{3.9}$$

And also we see that the same logic as (3.2) gives

$$P_{n,k} = P_{n-1,k}\left(1 - \frac{\lceil pk\rceil}{n}\right) + P_{n-1,k-1}\frac{\lceil p(k-1)\rceil}{n}. \tag{3.10}$$

Substituting (3.10) for $P_{n+1,k}$ in (3.9) yields

$$E(L_{n+1}^m) - E(L_n^m) = \sum_{k\ge1} k^m \frac{\lceil p(k-1)\rceil P_{n,k-1}}{n+1} - \sum_{k\ge1} k^m \frac{\lceil pk\rceil P_{n,k}}{n+1}, \tag{3.11}$$

and then we shift k by 1 in the first term (only), to obtain

$$E(L_{n+1}^m) - E(L_n^m) = \sum_{k\ge1} \frac{(k+1)^m\lceil pk\rceil P_{n,k} - k^m\lceil pk\rceil P_{n,k}}{n+1}$$
$$= \frac{E(((L_n+1)^m - L_n^m)\lceil pL_n\rceil)}{n+1}. \tag{3.12}$$

We sum equation (3.12) over all values of n and multiply by powers of z, to make generating functions. We obtain

$$\sum_{n\ge1} E(L_{n+1}^m)z^{n+1} - \sum_{n\ge1} E(L_n^m)z^{n+1} = \sum_{n\ge1} \frac{E((L_n+1)^m - L_n^m)\lceil pL_n\rceil)}{n+1}z^{n+1}.$$

We take a derivative, and perform some algebraic manipulations. This yields a differential equation that generalizes the one found in Gaither and Ward [3].

We first define

$$g(z) = \sum_{n\ge1} E(L_n^m)z^n,$$

and

$$f(z) = \sum_{n\ge1} E(((L_n+1)^m - L_n^m)\lceil pL_n\rceil - mpL_n^m)z^n.$$

The coefficient of the generating function $g(z)$ are our desired moments. The function $f(z)$ is designed so that the pole will be of the correct order later.

The algebraic manipulation mentioned above yield the nice differential equation

$$(1-z)g'(z) - 1 = (mp+1)g(z) + f(z), \tag{3.13}$$

which we can solve to obtain

$$g(z) = \frac{\int_0^z (1+f(t))(1-t)^{mp}dt}{(1-z)^{mp+1}}.$$

The function $g(z)$ has a singularity at $z = 1$ of order $mp+1$, which lines up with the form for the moment in equation (3.1), implying this to be the dominant pole. We then conclude from singularity analysis (see, for instance, [4]) that

$$c_{m,p} = \frac{\int_0^1 (1+f(t))(1-t)^{mp}dt}{\Gamma(mp+1)}. \tag{3.14}$$

We now perform a series of steps to simplify this form, following the methodology of the argument in [3]. We observe that

$$\int_0^1 (1-t)^{mp}dt = \frac{1}{mp+1},$$

and

$$c_{m,p} = \frac{1}{(mp+1)\Gamma(mp+1)} + \frac{\int_0^1 (f(t))(1-t)^{mp} dt}{\Gamma(mp+1)}.$$

We expand the numerator by substituting

$$\sum_{i\geq 1}\sum_{k\geq 1}(((k+1)^m - k^m)\lceil pk\rceil - mpk^m)P_{i,k}t^i$$

for $f(t)$, and using the Maclaurin series

$$(1-t)^{mp} = \sum_{n\geq 0}\frac{\Gamma(n-mp)}{\Gamma(n+1)\Gamma(-mp)}t^n,$$

allowing us to evaluate the integral. This yields

$$\begin{aligned} f(t)(1-t)^{mp} = \sum_{k\geq 1}&(((k+1)^m - k^m)\lceil pk\rceil - mpk^m) \\ &\times \sum_{i\geq 1} P_{i,k}\sum_{n\geq 0}\frac{\Gamma(n-mp)}{\Gamma(n+1)\Gamma(-mp)}t^{i+n} \end{aligned}$$

and

$$\begin{aligned} \int_0^1 f(t)(1-t)^{mp} = \sum_{k\geq 1}&((k+1)^m - k^m)\lceil pk\rceil - mpk^m) \\ &\times \sum_{i\geq 1} P_{i,k}\sum_{n\geq 0}\frac{\Gamma(n-mp)}{\Gamma(n+1)\Gamma(-mp)}\frac{1}{i+n+1}. \end{aligned}$$

In addition, we can simplify the above equation, using

$$\sum_{n\geq 0}\frac{\Gamma(n-mp)}{\Gamma(n+1)\Gamma(-mp)}\frac{1}{i+n+1} = \frac{i!\Gamma(mp+1)}{\Gamma(i+mp+2)}.$$

This yields

$$\begin{aligned} c_{m,p} = &\frac{1}{(mp+1)\Gamma(mp+1)} \\ &+ \sum_{k\geq 1}((k+1)^m - k^m)\lceil pk\rceil - mpk^m) \\ &\times \sum_{i\geq 1}\frac{i!P_{i,k}}{\Gamma(i+mp+2)}. \end{aligned}$$

Corollary 4 from [3] states that

$$\sum_{i\geq 1}\frac{i!P_{i,k}}{\Gamma(i+p+2)} = \frac{\lceil p(k-1)\rceil}{\lceil pk\rceil + p}\sum_{n\geq 1}\frac{n!P_{n,k-1}}{\Gamma(n+p+2)}.$$

We must modify this slightly, as follows

$$\begin{aligned} &\sum_{i\geq 1}\frac{i!P_{i,k}}{\Gamma(i+mp+2)} \\ &= \sum_{i\geq 1}\frac{i!\lceil p(k-1)\rceil\sum_{n<i}\frac{P_{n,k-1}}{n+1}\prod_{l=n+2}^{i}(1-\frac{\lceil pk\rceil}{l})}{\Gamma(i+mp+2)} \\ &= \lceil p(k-1)\rceil\sum_{n\geq 1}\frac{P_{n,k-1}}{n+1}\sum_{i>n}\frac{i!\prod_{l=n+2}^{i}(1-\frac{\lceil pk\rceil}{l})}{\Gamma(i+mp+2)} \\ &= \lceil p(k-1)\rceil\sum_{n\geq 1}\frac{P_{n,k-1}}{n+1}\frac{(n+1)!}{(\lceil pk\rceil + mp)\Gamma(n+mp+2)} \\ &= \frac{\lceil p(k-1)\rceil}{\lceil pk\rceil + mp}\sum_{n\geq 1}\frac{n!P_{n,k-1}}{\Gamma(n+mp+2)}. \end{aligned}$$

Substituting and applying $k-1$ times yields

$$\begin{aligned} c_{m,p} = &\frac{1}{(mp+1)\Gamma(mp+1)} \\ &+ \sum_{k\geq 1}((k+1)^m - k^m)\lceil pk\rceil - mpk^m) \\ &\times \prod_{j=2}^{k}(\frac{\lceil p(j-1)\rceil}{\lceil pj\rceil + mp})\sum_{i\geq 1}\frac{i!P_{i,1}}{\Gamma(i+mp+2)}. \end{aligned} \tag{3.15}$$

We have two simplifications. The first one is that

$$\begin{aligned} \prod_{j=2}^{k}(\frac{\lceil p(j-1)\rceil}{\lceil pj\rceil + mp}) &= \frac{1+mp}{\lceil pk\rceil}\prod_{j=1}^{k}(\frac{\lceil pj\rceil}{\lceil pj\rceil + mp}) \\ &= \frac{1+mp}{\lceil pk\rceil}\prod_{j=1}^{k}(\frac{1}{1+\frac{mp}{\lceil pj\rceil}}). \end{aligned}$$

As noted in [3], $P_{i,1}$ means that after i items have been seen, only one item has been retained. Since the first item is always retained, this retained item must be the first item, and thus this occurs with probability $1/i$. We then conclude

$$\begin{aligned} \sum_{i\geq 1}\frac{i!P_{i,1}}{\Gamma(i+mp+2)} &= \sum_{i\geq 1}\frac{(i-1)!}{\Gamma(i+mp+2)} \\ &= \frac{1}{(mp+1)^2\Gamma(mp+1)}. \end{aligned}$$

Combining these simplifications into (3.15), we obtain the form

$$c_{m,p} = \frac{1 + \sum_{k\geq 1}((k+1)^m - k^m - \frac{mpk^m}{\lceil pk\rceil})\prod_{j=1}^{k}(\frac{1}{1+\frac{mp}{\lceil pj\rceil}})}{(mp+1)\Gamma(mp+1)}.$$

Similarly, a form for $p = r/s$ is formed by the substitution $k = ls + b$, noting how the product cycles based on the remainder modulo s. We have

$$c_{m,p} = \frac{1}{(mp+1)\Gamma(mp+1)} \times \Big(1 + \sum_{(ls+b)\geq 1} \Big(((ls+b)+1)^m - (ls+b)^m - \frac{mp(ls+b)^m}{\lceil p(ls+b)\rceil}\Big) \prod_{j=1}^{(ls+b)} \Big(\frac{1}{1+\frac{mp}{\lceil pj\rceil}}\Big)\Big).$$

Mainly we see that $\lceil p(ls+b)\rceil = \lceil rl+pb\rceil = rl+\lceil pb\rceil$, and we break the product into l complete cycles through the s residues, times a remainder. We can then break the product in the following way to obtain the form given in main results.

$$\prod_{j=1}^{ls+b} \Big(\frac{1}{1+\frac{mp}{\lceil pj\rceil}}\Big) = \Big(\prod_{j=1}^{ls} \frac{1}{1+\frac{mp}{\lceil pj\rceil}}\Big)\Big(\prod_{i=ls+1}^{ls+b} \frac{1}{1+\frac{mp}{\lceil pi\rceil}}\Big)$$
$$= \Big(\prod_{\sigma=1}^{l}\prod_{j=1}^{s} \frac{1}{1+\frac{mp}{(\sigma-1)r+\lceil pj\rceil}}\Big) \times \Big(\prod_{i=1}^{b} \frac{1}{1+\frac{mp}{rl+\lceil pi\rceil}}\Big).$$

4 Future Results

After having analyzed all of the moments of L_n, a natural question is to character the distribution of L_n as well. We are almost finished this such a characterization, and we are likely to present the distributional results in our ANALCO seminar and/or in a future paper.

p	$c_{1,p}$	Decimal
1/1	$1/2$	0.5
1/2	$\frac{2}{3}\Gamma(1/2)$	1.1816
1/3	$\frac{1}{4}\Gamma(1/3)^2$	1.7942
2/3	$\frac{3}{10}2^{1/3}\Gamma(1/3)$	1.0125
1/4	$\frac{1}{20}\Gamma(1/4)^3$	2.3829
3/4	$\frac{4}{21}3^{1/4}\Gamma(1/4)$	0.9088
1/5	$\frac{1}{150}\Gamma(1/5)^4$	2.9613
2/5	$\frac{1}{14}2^{3/5}\frac{\Gamma(7/10)}{\Gamma(1/2)}\Gamma(1/5)^2$	1.6711
3/5	$\frac{5}{24}3^{2/5}\frac{\Gamma(13/15)}{\Gamma(2/3)}\Gamma(1/5)$	1.2025
4/5	$\frac{5}{36}4^{1/5}\Gamma(1/5)$	0.8413
1/6	$\frac{1}{1512}\Gamma(1/6)^5$	3.5342
5/6	$\frac{6}{55}5^{1/6}\Gamma(1/6)$	0.7941

p	$c_{2,p}$	Decimal
1/1	$1/3$	0.3333
1/2	2	2
1/3	$\frac{12}{5}\Gamma(2/3)^2$	4.4007
2/3	$\frac{9}{14}2^{2/3}\Gamma(2/3)$	1.3818
1/4	$\frac{4}{3}\Gamma(1/2)^3$	7.4244
3/4	$\frac{16}{45}3^{1/2}\Gamma(1/2)$	1.0915
1/5	$\frac{16}{35}\Gamma(2/5)^4$	11.0668
2/5	$\frac{5}{9}2^{6/5}\frac{\Gamma(9/10)}{\Gamma(1/2)}\Gamma(2/5)^2$	3.7862
3/5	$\frac{2}{891}3^{4/5}\frac{\Gamma(16/15)}{\Gamma(2/3)}\Gamma(2/5)$	1.9240
4/5	$\frac{25}{104}4^{2/5}\Gamma(2/5)$	0.9284
1/6	$\frac{1}{9}\Gamma(1/3)^5$	15.3310
5/6	$\frac{9}{50}5^{1/3}\Gamma(1/3)$	0.8246

p	$c_{3,p}$	Decimal
1/1	$1/4$	0.25
1/2	$\frac{24}{5}\Gamma(3/2)$	4.2539
1/3	$\frac{27}{2}$	13.5
2/3	$\frac{9}{4}$	2.25
1/4	$\frac{108}{7}\Gamma(3/4)^3$	28.3908
3/4	$\frac{64}{117}3^{3/4}\Gamma(3/4)$	1.5280
1/5	$\frac{81}{8}\Gamma(3/5)^4$	49.7964
2/5	$\frac{225}{88}2^{9/5}\frac{\Gamma(11/10)}{\Gamma(1/2)}\Gamma(3/5)^2$	10.5980
3/5	$\frac{40}{567}3^{6/5}\frac{\Gamma(19/15)}{\Gamma(2/3)}\Gamma(3/5)$	3.6824
4/5	$\frac{375}{1088}4^{3/5}\Gamma(3/5)$	1.1792
1/6	$\frac{9}{2}\Gamma(1/2)^5$	78.7204
5/6	$\frac{216}{875}5^{1/2}\Gamma(1/2)$	0.9784

References

[1] T. Ferguson. Who solved the secretary problem? *Statistical Science*, 4:282–296, 1989.

[2] A.M. Krieger, M. Pollak, and E. Samuel-Cahn. Select sets: Rank and file. *Annals of Applied Probability*, 17(1):360–385, 2007.

[3] J. Gaither and M.D. Ward. Analytic methods for select sets. *Probability in the Engineering and Informational Sciences*, 26(4):561–568, 2012.

[4] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

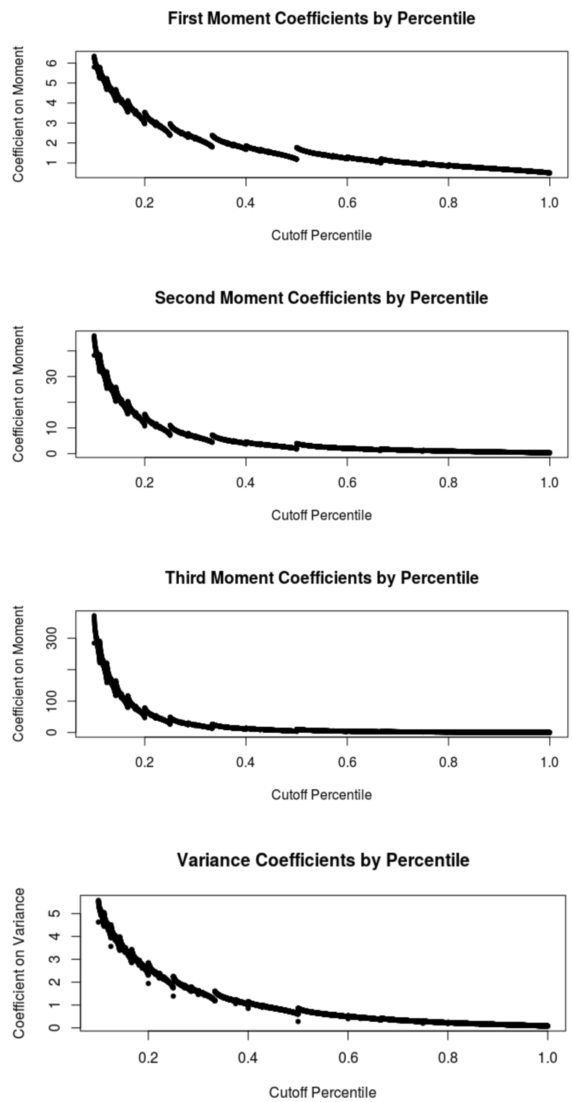

Figure 1: The first three graphs show the coefficients $c_{1,p}$, $c_{2,p}$, $c_{3,p}$, respectively, that are relevant to the asymptotics of the first, second, and third moments of the number of selected items. The fourth graph shows the coefficients for the variance.