Asymptotics for the Pareto Box Problem

Alexander Kovačec, Dep. Math. Univ. Coimbra, 3001-454 Coimbra, Portugal. kovacec@mat.uc.pt

Endow the unit n-cube $C = [0,1]^n$, with the componentwise order: $\underline{x} = (x_1, \ldots, x_n) \leq \underline{y} = (y_1, \ldots, y_n)$ iff $x_i \leq y_i$ for $i = 1, \ldots, n$. Then C is a partially ordered set and every finite nonempty subset M of C has a uniquely defined set of minimal elements, $P(M) = \min M$ called by mathematical economists sometimes the Pareto-set of M, I think.

Let us define a point p in C to be *random* if the coordinates of p are randomly selected with respect to the uniformly distribution in the interval $[0,1]$. A *random set* in C is one in whose points are random.

Mario Koeppen [K] has recently found

Theorem K. *Let $n, m \in \mathbb{Z}_{\geq 1}$. Consider a random set M of m points in $C = [0,1]^n$. Let the expected size of $P(M)$ be $e(m,n)$. Then:*
i. The $e(m,n)$ satisfy the recursion

$$e(m,1) = e(1,n) = 1, \textit{ and } e(m,n) = e(m-1,n) + \tfrac{1}{m}e(m,n-1), \textit{ for all } m,n \geq 2.$$

ii. An explicit formula for the $e(m,n)$ is given by

$$e(m,n) = \sum_{k=1}^{m} \frac{-^k}{k^{n-1}} \binom{m}{k}.$$

Koeppen derived his formula in ii from i by induction.

On these pages we give a method to asymptotically estimate for given n the function $\mathbb{Z}_{\geq 1} \ni m \mapsto e(m,n)$. We actually will use theorem Ki as a starting point and first give one more exact representation for $e(m,n)$.

Lemma 1. For $m, n \in \mathbb{Z}_{\geq 1}$ there holds

$$e(m,n) = \sum \{(j_1 j_2 \cdots j_{n-1})^{-1} : 1 \leq j_1 \leq \ldots \leq j_{n-1} \leq m\}$$

Proof. We use induction on n. For $n = 1$ the claim $e(m,1) = 1$ follows by obeying conventions concerning sums over empty products. For $n = 2$ the claim is that $e(m,2) = \sum_{j=1}^{m} \frac{1}{j}$. This holds true as is immediate from theorem Ki. Assuming the claim true for a fixed integer n, and introducing $e(0,n) := 0$, we can write

$$\begin{aligned}
e(m,n+1) &= \sum_{l=1}^{m} (e(l,n+1) - e(l-1,n+1)) \\
&= \sum_{l=1}^{m} \frac{1}{l} e(l,n) \\
&= \sum_{l=1}^{m} \frac{1}{l} \sum \{(j_1 j_2 \cdots j_{n-1})^{-1} : 1 \leq j_1 \leq \ldots \leq j_{n-1} \leq l\} \\
&= \sum_{l=1}^{m} \sum \{(j_1 j_2 \cdots j_{n-1} l)^{-1} : 1 \leq j_1 \leq \ldots \leq j_{n-1} \leq l\} \\
&= \sum \{(j_1 j_2 \cdots j_{n-1} l)^{-1} : 1 \leq j_1 \leq \ldots \leq j_{n-1} \leq l \leq m\} \\
&= \sum \{(j_1 j_2 \cdots j_{n-1} j_n)^{-1} : 1 \leq j_1 \leq \ldots \leq j_{n-1} \leq j_n \leq m\},
\end{aligned}$$

as we wished to show. □

The following definitions are standard.

Definition. Let $k \in \mathbb{Z}_{\geq 1}$, $\underline{x} = (x_1, \ldots, x_n)$. The *power sum* $p_k = p_k(\underline{x})$ and the *complete symmetric polynomial* $h_k = h_k(\underline{x})$ are defined respectively as

$$p_k(\underline{x}) = x_1^k + \ldots + x_n^k \quad \text{and} \quad h_k(\underline{x}) = \sum\{x_1^{l_1} x_2^{l_2} \cdots x_n^{l_n} : l_i \in \mathbb{Z}_{\geq 0},\, l_1 + \ldots + l_n = k\}; \text{ convention: } h_0 = 1.$$

So p_k and h_k are homogeneous and symmetric polynomials of degree k. Note that $h_k(\underline{x})$ can be alternatively presented as $h_k(\underline{x}) = \sum\{x_{j_1} x_{j_2} \cdots x_{j_k} : 1 \leq j_1 \leq \ldots \leq j_k \leq n\}$.

It follows from this representation that $e(m, n)$ can actually be viewn as the value of the complete symmetric polynomial of degree $n - 1$ in m variables in the point $(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{m})$: $e(m, n) = h_{n-1}(1, \frac{1}{2}, \ldots, \frac{1}{m})$.

An *ordered partition* of an integer n is a decreasing uple $\lambda = (\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k)$ of integers≥ 1, of sum n: $\lambda_k \geq 1$, $\sum_i \lambda_i = n$. Convention: for $i > k$, $\lambda_i = 0$. One frequently writes $\lambda \vdash n$ to indicate these two properties. Integer $k = l(\lambda)$ is the *length* of λ. Also, λ_1-uple $\underline{m} = (m_1(\lambda), ..., m_{\lambda_1}(\lambda))$ is given via $m_i(\lambda)$ =number of entries of λ equal to i. Convention: for $i > \lambda_1$, $m_i(\lambda) = 0$.

For a partition λ, one (see [M]) defines the egyptian fraction $z_\lambda = (\prod_{i \geq 1} m_i(\lambda)! \cdot i^{m_i(\lambda)})^{-1}$ and extends the definition of p_k via the definition $p_\lambda = p_{\lambda_1} p_{\lambda_2} \ldots p_{\lambda_k}$.

Now, for the complete symmetric polynomial of degree n, h_n the following is known.

Lemma 2. Whatever the number of variables comprising $\underline{x}$, polynomials $h_k(\underline{x})$ and $p_k(\underline{x})$ are related by the following two formulae:

$$k h_k = \sum_{r=1}^{k} p_r h_{k-r}, \qquad \text{and} \qquad h_k = \sum_{\lambda \vdash k} z_\lambda p_\lambda = \sum_{\lambda \vdash k} \left(\prod_{i \geq 1} m_i(\lambda)! i^{m_i(\lambda)} \right)^{-1} p_\lambda.$$

Proof. See [M, p24]. □

Example 3. a. The following table shows the seven ordered partitions λ of the integer 5, the associated uple $\underline{m}(\lambda)$, the egyptian fraction z_λ, and the associated summand $z_\lambda p_\lambda$.

λ	$\underline{m}(\lambda)$	z_λ	p_λ	$z_\lambda p_\lambda$
(5)	$(0,0,0,0,1)$	$1/5$	$\sum x_i^5$	$1/5 \sum x_i^5$
$(4,1)$	$(1,0,0,1)$	$1/4$	$\sum x_i^4 \cdot \sum x_i$	$1/4(\sum x_i^4)(\sum x_i)$
$(3,2)$	$(0,1,1)$	$1/6$	$\sum x_i^3 \cdot \sum x_i^2$	$1/6(\sum x_i^3)(\sum x_i^2)$
$(3,1,1)$	$(2,0,1)$	$1/6$	$\sum x_i^3 \cdot \sum x_i \cdot \sum x_i$	$1/6(\sum x_i^3)(\sum x_i)^2$
$(2,2,1)$	$(1,2)$	$1/8$	$\sum x_i^2 \cdot \sum x_i^2 \cdot \sum x_i$	$1/8(\sum x_i^2)^2(\sum x_i)$
$(2,1,1,1)$	$(3,1)$	$1/12$	$\sum x_i^2 \cdot \sum x_i \cdot \sum x_i \cdot \sum x_i$	$1/12(\sum x_i^2)(\sum x_i)^3$
$(1,1,1,1,1)$	(5)	$1/120$	$\sum x_i \cdot \sum x_i \cdot \sum x_i \cdot \sum x_i \cdot \sum x_i$	$1/120(\sum x_i)^5$

b. From our previous observations we conclude that

$$\begin{aligned} e(m,6) &= h_5(1, \frac{1}{2}, \ldots, \frac{1}{m}) \\ &= \frac{1}{5}\sum_{l=1}^{m}\frac{1}{l^5} + \frac{1}{4}\sum_{l=1}^{m}\frac{1}{l^4}\sum_{l=1}^{m}\frac{1}{l} + \frac{1}{6}\sum_{l=1}^{m}\frac{1}{l^3}\sum_{l=1}^{m}\frac{1}{l^2} + \frac{1}{6}(\sum_{l=1}^{m}\frac{1}{l^3})(\sum_{l=1}^{m}\frac{1}{l})^2 + \frac{1}{8}(\sum_{l=1}^{m}\frac{1}{l^2})^2(\sum_{l=1}^{m}\frac{1}{l}) + \\ &\quad + \frac{1}{12}(\sum_{l=1}^{m}\frac{1}{l^2})(\sum_{l=1}^{m}\frac{1}{l})^3 + \frac{1}{120}(\sum_{l=1}^{m}\frac{1}{l})^5 \qquad (1) \end{aligned}$$

In the following lemma, $\gamma \approx 0.5772...$ denotes Euler's constant, B_j the $j-$th Bernoulli number, and $\zeta(.)$ Riemann's Zeta function.

Lemma 4. There hold the following asymptotic formulae

$$\sum_{l=1}^{m} \frac{1}{l} = \ln m + \gamma + \frac{1}{2m} - \frac{1}{12m^2} + \frac{\epsilon_m}{120m^4}, \quad \text{with} \quad 0 < \epsilon_m < 1.$$

b. For $m \in \mathbb{Z}_{\geq 1}$, $s \in \mathbb{R}_{\geq 2}$,

$$\sum_{l=1}^{m} \frac{1}{l^s} = \zeta(s) - \frac{1}{(s-1)m^{s-1}} - \frac{1}{2m^s} + \sum_{k=1}^{n} \frac{B_{2k}}{2k} \binom{-s}{2k-1} m^{-s-2k+1} + O(m^{-2-2k-1}).$$

Proof. See [GKP, p263c10, p477c8, 569c9] □

These formulae more than suffice to give an $o(m)$ estimate for $e(m,6)$.

Corollary 5. For $e(m,6) = e_{\text{appr1}}(m,6) + O((\log m)^4/m)$, where
$$e_{\text{appr1}}(m,6) = c_5(\ln m)^5 + c_4(\ln m)^4 + c_3(\ln m)^3 + c_2(\ln m)^2 + c_1(\ln m)^1 + c_0,$$

with

$$\begin{aligned}
c_5 &= \frac{1}{120} = 0.00833333333333333...,\\
c_4 &= \frac{\gamma}{24} = 0.0240506527042305...,\\
c_3 &= \frac{\gamma^2}{12} + \frac{\pi^2}{72} = 0.164842665887995...,\\
c_2 &= \frac{\gamma^3}{12} + \frac{\gamma\pi^2}{24} + \frac{\zeta(3)}{6} = 0.453739538040443...,\\
c_1 &= \frac{\gamma^4}{24} + \frac{\gamma^2\pi^2}{24} + \frac{\pi^4}{160} + \frac{\gamma\zeta(3)}{3} = 0.981728086834400...,\\
c_0 &= \frac{\gamma^5}{120} + \frac{\gamma^3\pi^2}{72} + \frac{\gamma\pi^4}{160} + \frac{\gamma^2\zeta(3)}{6} + \frac{\pi^2\zeta(3)}{36} + \frac{\zeta(5)}{5} = 0.981995068903145....
\end{aligned}$$

Proof. From lemma 4 we have the estimates
$$\sum_{l=1}^{m} \frac{1}{l} = \ln m + \gamma + O(\frac{1}{m}), \text{ and } \sum_{l=1}^{m} \frac{1}{l^k} = \zeta(k) + O(\frac{1}{m}), \text{ for } k \in \mathbb{Z}_{\geq 2}.$$

If we plug these estimates into the formula (1), and use the rules for computation with $O(*)$, and the numerical values for $\zeta(2), ..., \zeta(5)$ (in some of which π and the Bernoulli numbers creep in) and γ, the claim follows. □

To give an idea of the goodness of these estimates we give the following table; for the significance of function $e_{\text{appr2}}(m,6)$, see below.

Table 7.

m	1	2	3	4	5	6	7	8	9
$e(m,6)$	1	1.968	2.9103	3.82798	4.72409	5.6007	6.45948	7.30186	8.12903
$e_{\text{appr1}}(m,6)$	0.98199	1.94226	2.87512	3.78563	4.67592	5.54769	6.40242	7.24135	8.06557
$e_{\text{appr2}}(m,6)$	0.999998	1.96803	2.91022	3.82805	4.72412	5.60064	6.4594	7.30187	8.12927
m	10	20	30	50	100	200	2000	20000	1000000
$e(m,6)$	8.94206	16.453	23.1509	34.98	59.38	97.2557	?	?	?
$e_{\text{appr1}}(m,6)$	8.87605	16.3745	23.0672	34.90	59.30	97.18	398.742	1240.57	5606.25
$e_{\text{appr2}}(m,6)$	8.94266	16.4644	23.1778	35.0548	59.5518	97.6144	400.933	1248.15	5641.06

The reader may notice that the absolute error of $e_{\text{appr1}}(m,6)$ is slightly increasing till somewhere between 50 and 100, and then seems to fall again. The reason is that our estimate is of order $O((\log n)^4/n)$ and the function $(\log n)^4/n$ peaks at $n=55$. The determination of the precise value of $e(m,6)$ for $m \geq$several hundred seems to be out of question for older PCs (at least) and Mathematica. □

The first part of the corollary is a special case of a theorem that allows another approach to find asymptotic estimates for $m \mapsto e(m,n)$ with good results.

Theorem 6. *Given $n \in \mathbb{Z}_{\geq 1}$, there exist real constants c_j, $j=0,1,2,\dots,n-1$, such that*

$$e_{\text{asym}}(m,n) = \frac{1}{(n-1)!}(\ln m)^{n-1} + \sum_{j=n-2:-1:0} c_j (\ln m)^j$$

is an $O((\log m)^{n-2}/m)$ approximation to $e(m,n)$.

Proof. For $\lambda \vdash k$, we can write, $p_\lambda = \prod_{i=1}^k p_i^{m_i(\lambda)}$. Consequently, generalizing formula (1) of lemma 2,

$$e(m,n) = h_{n-1}(1,\dots,\frac{1}{m}) = \sum_{\lambda \vdash n-1} z_\lambda p_\lambda = \sum_{\lambda \vdash n-1} \prod_{i \geq 1} \left((m_i(\lambda)! i^{m_i(\lambda)})^{-1} (\sum_{j=1}^{m} \frac{1}{j^i})^{m_i(\lambda)} \right).$$

If we now substitute the estimates given in the proof of corollary 5, we get, analogously as there, the claim. □

This result suggests the following two-step approach to find an asymptotic estimate for $e(m,n)$.

1. Compute the precise values of $e(m,n)$ for a set D of small ms using theorem Kii.
2. Find c_js, $j=0,1,2,\dots,n-2$, by a least square procedure; that is, minimize $\sum_{m \in D} |e(m,n) - e_{\text{appr}}(m,n)|^2$.

That this procedure seems to yield good results again is shown in the following example.

Example 8. If we take from Table 7 the values $e(m,6) - \frac{1}{120}\ln(m)^5$ for $m \in D = \{1,2,3,4,5,6,8\}$, we get the following table suitable for a Mathematica©-input; see [W, p672,p539]

`data1={{1,1},{2,1.96667},{3,2.89696},{4,3.78531},{5,4.6341},{6,5.44681},{8,6.97785}}.`

The list of functions whose linear combination should be a least-square-fit through these data points is `funs={1, Log[m],Log[m]^2, Log[m]^3, Log[m]^4 }`. Mathematica© uses `Log[m]` for the natural logarithm. Now typing `N[Fit[data1, funs, m]+(1/120)*Log[m]^5]` yields as a substitute for the academic asymptotic e_{asym} of theorem 6 the practical approximation

$$e_{\text{appr2}}(m,6) = 0.08333(\ln m)^5 + 0.02561*(\ln m)^4 + 0.15462(\ln m)^3 + 0.47898(\ln m)^2 + 0.97982(\ln m) + 0.99999,$$

where the decimal places where chopped off after the 5th. With this function, one produces the third lines of the table 7. One sees that near to points of the set D the approximation is predictably even better than $e_{\text{appr1}}(m,6)$, but for most problems, $e_{\text{appr2}}(m,n)$, for large n easier to compute than $e_{\text{appr1}}(m,n)$, will also behave sufficiently well if m is large and far outside the set D choosen for its definition. However, for $o(m)$-estimates one has to use e_{asym} and in practice to augment the precision of the constants c_j as occuring in theorem 6 correspondingly.

References

[M] I. G. Macdonald, Symmetric Functions and Hall Polynomials, Clarendon, Oxford, 1995.

[GKP] Graham-Knuth-Patashnik, Concrete Mathematics, Addison Wesley 1989.

[K] M. Koeppen, Analysis of the Pareto Box Problem. Preprint, Fraunhofer IPK, Berlin 2005.

[W] S. Wolfram, Mathematica©, A System for doing Mathematics by Computer, Addison-Wesley 1991 (much later editions available).