Extremes of Pitman's Random Partition and their Asymptotics

Shuhei Mano

The Institite of Statistical Mathematics

Satellite Meeting to ISI Congress 2013
@ Sanjo Hall, University of Tokyo
September 4, 2013

Poisson-Dirichlet Distribution

Definition (Kingman 1978; Pitman 1995)

For $0 \le \alpha < 1$ and $\theta > -\alpha$, or $\alpha < 0$ and $\theta = -\alpha m$, $m \in \mathbb{N}$, let

$$P_1 = W_1, \qquad P_i = W_i \prod_{j=1}^{i-1}(1 - W_j), \qquad i = 2, 3, \ldots,$$

with $W_i \sim \text{Beta}(1-\alpha, \theta + i\alpha)$, iid. The ranked sequence of (P_i) is the 2-parameter Poisson-Dirichlet distribution $PD(\alpha, \theta)$.

Remark

- $P(\cdot) = \sum_{i=1}^{\infty} P_i \delta_{Y_i}(\cdot)$ is the 2-parameter Dirichlet process.
- Most general such that (P_i) is invariant under size-biased permutation.

Pitman's Random Partition

A partition of $\mathbf{n} \in \mathbb{N}$ by integers is identified by multiplicity $(\mathbf{c_i})$ such that

$$\|\mathbf{c}\| := \sum_{i=1}^{n} c_i = k_n, \qquad |\mathbf{c}| := \sum_{i=1}^{n} i c_i = n.$$

Definition (Ewens 1972; Pitman 1992)

An exchangeable random partition

$$P((C_i) = (c_i), K_n = k) = \frac{\left(\frac{\theta}{\alpha}\right)_k}{(\theta)_n} (-1)^{n-k} n! \prod_{j=1}^{n} \binom{\alpha}{j}^{c_j} \frac{1}{c_j!},$$

where $(\theta)_n = \theta(\theta+1)\cdots(\theta+n-1)$.

The sampling distribution from $\mathbf{PD}(\alpha, \theta)$; the ranked sizes converges as

$$n^{-1}(L_1^{(n)}, L_2^{(n)}, \ldots) \xrightarrow{d} (P_{(1)}, P_{(2)}, \ldots), \qquad n \to \infty,$$

but how are the limiting distributions and in $\mathbf{o(n)}$?

Incomplete Dirichlet Integrals

A collection of random variables $\mathbf{(Y_1, ..., Y_d) \sim Dir(\nu; \rho)}$ if the pdf is

$$\frac{\Gamma(\rho + d\nu)}{\Gamma(\rho)\Gamma(\nu)^d} \left(1 - \sum_{j=1}^{d} y_j\right)^{\rho-1} \prod_{i=1}^{d} y_i^{\nu-1}$$

over the simplex $\Delta_d = \{0 < y_i, i = 1, ..., d, \sum_{j=1}^{d} y_j < 1\}$. Let us introduce an extension of the incomplete Dirichlet integrals of Type I (Sobel et al. 1970)

$$\mathcal{I}_{p,q}^{(d)}(\nu; \rho) := \frac{\Gamma(\rho + d\nu)}{\Gamma(\rho)\Gamma(\nu)^d} \int_{\Delta_d(p,q)} \left(1 - \sum_{j=1}^{d} y_j\right)^{\rho-1} \prod_{i=1}^{d} y_i^{\nu-1} dy_i,$$

where $\Delta_d(p, q) = \{p < y_i, i = 1, ..., d; \sum_{j=1}^{d} y_j < 1 - q\}$.

Limiting Distribution of Maximum

For $\alpha < 0$ and $\theta = -\alpha m$, $m \in \mathbb{N}$, $(P_i) \sim \mathrm{Dir}(-\alpha)$, so we skip and assume $\alpha \geq 0$.

Theorem (Pitman & Yor 1997)

For $0 \leq \alpha < 1$ *and* $\theta > -\alpha$*, the maximium,* $P_{(1)}$*, has cdf*

$$\rho_{\alpha,\theta}(x) = 1 + \sum_{k=1}^{\lfloor x^{-1} \rfloor} \frac{\alpha^{-k}(\theta)_{k:\alpha}}{k!} \mathcal{I}_{x,0}^{(k)}(-\alpha; k\alpha + \theta),$$

where $(\theta)_{k:\alpha} = \theta(\theta + \alpha) \cdots (\theta + (k-1)\alpha)$*.*

Remark

- Appears as the limiting distribution of $n^{-1} L_1^{(n)}$ with $r/n \to x$.
- In the number theory $\rho_{0,1}(u)$ is known as Dickman's function (1930) on the distribution of the smooth numbers.

Limiting Distribution of Minimum

Theorem (Arratia, Barbour, Tavaré 2003)

For $\alpha = 0$ *and* $\theta > 0$, *the minimum*, $\mathbf{L}^{(n)}_{K_n}$, *has*

$$\mathbf{P}(\mathbf{L}^{(n)}_{K_n} > r) \sim \Gamma(\theta)(nx)^{-\theta}\omega_\theta(x), \qquad n, r \to \infty, \qquad r/n \to x,$$

where

$$\omega_\theta(x) = \theta x^\theta \left\{ 1 + \sum_{1 \le k \le x^{-1} - 1} \frac{\theta^k}{(k+1)!} \mathcal{I}^{(k)}_{x,x}(0;0) \right\}.$$

Remark

The distribution is degenerate. In the number theory $\omega_1(x)$ is known as Buchstab's function (1937) on the distribution of the rough numbers.

Limiting Distribution of Minimum (cont.)

Theorem (Arratia, Tavaré 1992)

For $\alpha = 0$ and $\theta > 0$, the minimum, $L_{K_n}^{(n)}$, has

$$P(L_{K_n}^{(n)} > r) \sim \exp\left(-\theta \sum_{j=1}^{r-1} j^{-1}\right), \qquad n \to \infty, \qquad r = o(n).$$

For $0 < \alpha < 1$ and $\theta > -\alpha$, the limiting distribution has not been considered, but $L_{k_n}^{(n)} \xrightarrow{p} 1$, since

Theorem (Yamato & Sibuya 2000)

$$n^{-\alpha}(C_1, ..., C_m) \xrightarrow{d} (p_\alpha(1), ..., p_\alpha(m))M,\ i = 1, 2, ..., m,\ n \to \infty,$$

where $p_\alpha(i) = \binom{\alpha}{i}(-1)^{i+1}$ *is pmf of Sibuya's distribution (1979)*

and M *has pdf of* $x^{\theta/\alpha} \times$ *Mittag-Leffler's distribution.*

Outline

In summary, known facts are

	$P(L_1^{(n)} < r)$		$P(L_{K_n}^{(n)} > r)$	
r	**O(n)**	**o(n)**	**O(n)**	**o(n)**
$\alpha < 0$		$\mathbf{Dir}(-\alpha)$ and back to Fisher (1929)		
$\alpha = 0$	Dickman[1]	?	Buchstab[2] (degenerate)	Poisson[3]
$\alpha > 0$	Dickman[1]	?	? (degenerate[4])	? (degenerate[4])

1 Pitman & Yor (1997); 2 Arratia et al (2003); 3 Arratia & Tavaré (1992); 4 Yamato & Sibuya (2000). But for the random permutation ($\alpha = 0$, $\theta = 1$) it is a classic back to Goncharov (1942), Shepp & Lloyd (1966).

In this talk, the table is filled out with reproductions. All results have been obtained by probability techniques, but here singularity analysis in analytic combinatrics (Flajolet & Odolyzko 1990) is employed.

Use of Sufficiency

Assume α is known.

$$\begin{aligned} P(K_n = k) &= \sum_{\|c\|=k} P((C_i) = (c_i), K_n = k) \\ &= \frac{\left(\frac{\theta}{\alpha}\right)_k}{(\theta)_n} (-1)^{n-k} n! \sum_{\|c\|=k} \prod_{j=1}^{n} \left(\begin{array}{c} \alpha \\ j \end{array} \right)^{c_j} \frac{1}{c_j!} \\ &= \frac{\left(\frac{\theta}{\alpha}\right)_k}{(\theta)_n} (-1)^{n-k} C(n, k; \alpha). \end{aligned}$$

K_n is the complete and sufficient statistic of θ:

$$P((C_i) = (c_i) | K_n = k) = \frac{n!}{C(n, k; \alpha)} \prod_{j=1}^{n} \left(\begin{array}{c} \alpha \\ j \end{array} \right)^{c_j} \frac{1}{c_j!},$$

where $C(n, k; \alpha)$ is the generalized factorial coefficient .

Generalized Factorial Coefficient

A generalized factorial of $\mathbf{t}$ of order $\mathbf{n}$ and scale parameter is α,

$$[\alpha t]_n = \alpha t(\alpha t - 1)\cdots(\alpha t - n + 1), \qquad [\alpha t]_0 = 1.$$

The generalized factorial coefficients is defined as (Charalambides 2005)

$$[\alpha t]_n = \sum_{k=1}^{n} C(n, k; \alpha)[t]_k,$$

with the exponential generating function (egf):

$$\sum_{n=k}^{\infty} C(n, k; \alpha)\frac{u^n}{n!} = \frac{1}{k!}\left((1+u)^{\alpha} - 1\right)^k.$$

Generalized Factorial Coefficient (cont.)

Suppose $\alpha\,(\geq \mathbf{n}) \in \mathbb{N}$ and partitions of **n** like balls into **k** urns. Each urn consists of α distinguishable cells and each cell accept one ball.

$$\sum_{|\mathbf{a}|=\mathbf{n}} \prod_{\mathbf{j}=1}^{\mathbf{k}} \binom{\alpha}{\mathbf{a_j}}.$$

Let $\mathbf{c_i = |\{j : a_j = i\}|}$.

$$\sum_{\|\mathbf{c}\|=\mathbf{k}} \frac{\mathbf{k!}}{\mathbf{c_1!c_2!\cdots c_n!}} \prod_{\mathbf{j}=1}^{\mathbf{n}} \binom{\alpha}{\mathbf{j}}^{\mathbf{c_j}} = \frac{\mathbf{k!}}{\mathbf{n!}}\mathbf{C(n, k; \alpha)}.$$

Example ($\mathbf{K_4 = 2}$)

$(1,2)(1,2)$	$[\alpha]_2^2/(2!)^2$
$(1)(1,2,3)$,$(1,2,3)(1)$	$2\alpha[\alpha]_3/(1!\cdot 3!)$

Therefore $\mathbf{C(4,2;\alpha) = 3[\alpha]_2^2 + 4\alpha[\alpha]_3}$.

Associated Generalized Factorial Coefficients

If each urn has at least $\mathbf{r}$ balls,

$$\mathbf{n!}\sum_{c_{j<r}=0}\prod_{j=r}^{n}\binom{\alpha}{j}^{c_j}\frac{1}{c_j!} =: C_r(n,k;\alpha), \qquad n = rk, rk+1, \ldots$$

which is known as the $\mathbf{r}$-associated generalized factorial coefficient (Charalambides 2005). Let us define an extension, in which (c_i) are restricted such that the $\mathbf{i}$-th smallest has at least $\mathbf{r}$ balls.

Definition (Mano 2013)

An extension of the $\mathbf{r}$-associated generalized factorial coefficient is

$$C_r^{(i)}(n,k;\alpha) := n!\sum_{\sum_{j=1}^{r-1}c_j<i}\prod_{j=r}^{n}\binom{\alpha}{j}^{c_j}\frac{1}{c_j!}, \qquad i = 1, \ldots, k,$$

where $n = rk-(i-1)(r-1), rk-(i-1)(r-1)+1, \ldots$.

Exact Marginal Distribution

Definition (cont.)

$$\mathbf{C}^{r(i)}(n,k;\alpha) := n! \sum_{\sum_{j=r+1}^{n} c_j < i} \prod_{j=r}^{n} \binom{\alpha}{j}^{c_j} \frac{1}{c_j!}, \qquad i = 1, \dots, k,$$

where $n = k, \dots, rk$ for $i = 1$ and $n = k, k+1, \dots$ for $i > 1$.

Lemma

The $\mathbf{i}$*-th smallest,* $L^{(n)}_{K_n - i + 1}$*, has*

$$P(L^{(n)}_{K_n-i+1} \geq r) = \sum_{k=1}^{\lfloor \frac{n+(i-1)(r-1)}{r} \rfloor} \frac{(-1)^n}{(-\alpha)^k} \frac{(\theta)_{k;\alpha}}{(\theta)_n} C^{(i)}_r(n,k;\alpha).$$

and the $\mathbf{i}$*-th largest,* $L^{(n)}_i$*, has cdf*

$$P(L^{(n)}_i \leq r) = \sum_{k=\lceil \frac{n}{r} \rceil, 1}^{n} \frac{(-1)^n}{(-\alpha)^k} \frac{(\theta)_{k;\alpha}}{(\theta)_n} C^{r(i)}(n,k;\alpha).$$

Exponential Generating Functions

Lemma

$$\sum_{n=k}^{rk} C^r(n,k;\alpha)\frac{u^n}{n!} = \frac{1}{k!}\left(\sum_{j=1}^{r}\binom{\alpha}{j}u^i\right)^k$$

and

$$\sum_{n=rk}^{\infty} C_r(n,k;\alpha)\frac{u^n}{n!} = \frac{1}{k!}\left((1+u)^\alpha - \sum_{j=0}^{r-1}\binom{\alpha}{j}u^j\right)^k.$$

Proof.

By combinatrics. But more general assertion is possible via Faà di Bruno's formula (1855; but Arbogast 1800; Comtet 1974; Shimizu et al. 2000); for $h(u) = g(f(u))$,

$$h^{(n)}(u) = \sum_{d=1}^{n} g^{(d)}(f(u)) \sum_{||c||=d} n! \prod_{j=1}^{n}\left(\frac{f^{(j)}(u)}{j!}\right)^{c_j}\frac{1}{c_j!},$$

Inversion Formulae

Proposition (Shimizu in private communication)

If an egf a $\mathbf{r}$*-associated number* $\mathbf{G_r(n,k)}$ *is defined in terms of a series* $\mathbf{q_j}$*,*

$$\sum_{n=rk}^{\infty} G_r(n,k)\frac{u^n}{n!} = \frac{1}{k!}\left(\sum_{j=r}^{\infty} q_j u^j\right)^k .$$

Then the number has an expression

$$G_r(n,k) = n!\sum_{c_{j<r}=0}\prod_{j=r}^{n}\frac{q_j^{c_j}}{c_j!}, \qquad n = rk, rk+1, \ldots$$

Remark

Similar formula holds for $\mathbf{G^r(n,k)}$. $\mathbf{j^{-1}}$ gives the $\mathbf{r}$-associated Stirling number of the first kind, $\binom{\alpha}{j}$ gives the $\mathbf{r}$-associated generalized factorial coefficient.

Minimum with $\alpha > 0$ and $r = o(n)$

Theorem

For $0 < \alpha < 1$ and $\theta > \alpha$, the minimum, $L_{K_n}^{(n)}$, has

$$P(L_{K_n} > r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha}(c_\alpha(r-1))^{-1-\frac{\theta}{\alpha}}, \; n \to \infty, \; r = o(n),$$

where $r = 2, 3, \ldots$ and $c_\alpha(r)$ is cdf of Sibuya's distribution.

Proof.

Applying the Cauchy-Goursat theorem to the egf,

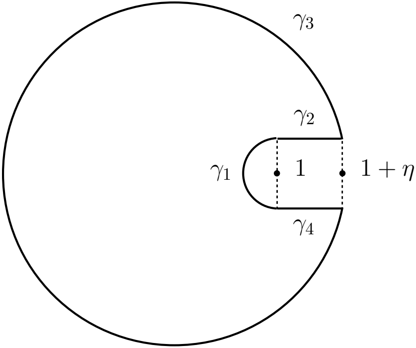
$$P(L_{K_n} > r) = \frac{n!}{[\theta]_n} \frac{1}{2\pi\sqrt{-1}} \oint \frac{(f_{\alpha,r}(u))^{-\frac{\theta}{\alpha}}}{u^{n+1}} du,$$

where $f_{\alpha,r}(u) = (1-u)^\alpha - \sum_{j=1}^{r-1} \binom{\alpha}{j} (-u)^j$.

Minimum with $\alpha > \mathbf{0}$ and $\mathbf{r} = \mathbf{o(n)}$ (cont.)

Proof. (cont.)

According to Rouché's theorem, we can show that $\mathbf{f_{\alpha,r}(u) = 0}$ has no roots in $|\mathbf{u}| \leq \mathbf{1}$. Therefore taking the cut $\mathbf{[1, \infty)}$ and $\boldsymbol{\eta} > \mathbf{0}$ we can deform the contour for the Cauchy integral without changing the value.



$\boldsymbol{\gamma_2}$, $\boldsymbol{\gamma_1}$, $\boldsymbol{\gamma_4}$ avoid the cut with distance $\mathbf{1/n}$. Contribution comes from the part of the contour, which is similar to Hankel's contour for the asymptotic expansion of the Gamma function. Taking $\mathbf{n} \rightarrow \infty$ the theorem follows. □

Maximum with $\alpha = 0$ and $r = o(n)$

Theorem

For $\alpha = 0$ and $\theta > 0$, the maximum, $L_1^{(n)}$, has cdf

$$P(L_1^{(n)} < r) \sim \frac{\Gamma(\theta) n^{-\theta+1/2}}{\sqrt{2\pi r}} \sum_{j=0}^{r-1} \rho_{\theta,r,n,j}^{-n} \exp\left(\theta \sum_{k=1}^{r} \frac{\rho_{\theta,r,n,j}^{k}}{k}\right), \; n \to \infty,$$

where $r = o(n)$ and $\rho_{\theta,r,n.j} \sim (n/\theta)^{1/r}$, $j = 0, 1, ..., r-1$ are the roots of the equation $u + u^2 + \cdots + u^r = (n+1)/\theta$.

Proof.

Applying the Cauchy-Goursat theorem to egf, we have an integral expression. Taking contour as a polygon which goes through each saddle point (absolute values are $\rho_{\theta,r,n.j}$) along the direction of the steepest descent, the Cauchy integral is evaluated. □

Summary

- Exact formule for marginal distributions of ranked sequence were obtained in terms of an extension of the associated generalized factorial coefficient.
- For the limiting distributions, singularity analysis yielded

	$P(L_1^{(n)} < r)$		$P(L_{K_n}^{(n)} > r)$	
r	$O(n)$	$o(n)$	$O(n)$	$o(n)$
$\alpha < 0$	Dickman	0	$\mathcal{I}_{x,x}^{(m-1)}(-\alpha, -\alpha)$	$1 - a_r n^{\alpha}$
$\alpha = 0$	Dickman	$b_r n^{1/2-\theta} \rho_r^{-n}$	Buchstab$\cdot n^{-\theta}$	Poisson
$\alpha > 0$	Dickman	$c_r n^{\theta/\alpha - \theta} \rho_r^{-n}$	$n^{-\theta-\alpha}$	Sibuya$\cdot n^{-\theta-\alpha}$

- Details: arXiv:1306.2056