The Poisson-Dirichlet distribution:

An approach from the theory of point processes *

Kenji Handa
(Saga University)

Subjects

Poisson-Dirichlet distributions [Kingman,1975]
Limit distributions of ordered statistics under Dirichlet laws
The two-parameter generalization [Pitman-Yor,1997]
The law of ranked values $V_1 > V_2 > \cdots$ of a 'residual allocation model' with ratios governed by independent beta laws

Contents

Results for P-D distributions
$\Leftarrow$ **analysis of the point process $\xi := \sum \delta_{V_i}$ on $(0,1]$**

*cf. Bernoulli 15(4) (2009) 1082-1116

Definition and a basic property

Residual allocation model $0 \le \alpha < 1, \theta > -\alpha$: **given**

$\{W_i\}_{i=1}^{\infty}$ **: independent,** $W_i \sim \mathrm{Beta}(1-\alpha, \theta+\alpha i)$

"stick breaking" $\begin{cases} \widetilde{V_1} := W_1 \\ \widetilde{V_2} := (1-W_1)W_2 \\ \widetilde{V_3} := (1-W_1)(1-W_2)W_3 \\ \cdots \end{cases}$

GEM(α,θ)**: the law of** $(\widetilde{V_1}, \widetilde{V_2}, \ldots)$

PD(α,θ)**: the law of the ranked sequence** $(V_1, V_2, \ldots)$ **of** $(\widetilde{V_1}, \widetilde{V_2}, \ldots)$

Fact **[Pitman, 1996] If** $(V_1, V_2, \ldots) \sim \mathrm{PD}(\alpha,\theta)$**, then** $(\widetilde{V_1}, \widetilde{V_2}, \ldots) \overset{\text{law}}{=} (V_{N_1}, V_{N_2}, \ldots)$**, where** $(N_i)_{i=1}^{\infty}$ **is the size-biased permutation, i.e.,** $P(N_1 = n \mid (V_i)_{i=1}^{\infty}) = V_n$ **and for** $j = 2, 3, \ldots$

$$P\left(N_j = n \,\middle|\, (V_i)_{i=1}^{\infty}, N_1, \ldots, N_{j-1}\right) = \frac{V_n \mathbf{1}_{\{n \neq N_1, \ldots, N_{j-1}\}}}{\sum_i V_i \mathbf{1}_{\{i \neq N_1, \ldots, N_{j-1}\}}}.$$

Correlation functions

For $(V_1, V_2, \ldots) \sim \mathrm{PD}(\alpha, \theta)$**, consider** $\xi = \sum_{i=1}^{\infty} \delta_{V_i} = \sum_{i=1}^{\infty} \delta_{\widetilde{V}_i}$**.**

Theorem 1 **The** n**th correlation functions** $q_{n,\alpha,\theta}$ **of** ξ**, i.e.,** $q_{n,\alpha,\theta} : \mathbf{R}^n \to [0, \infty)$ **such that**

$$q_{n,\alpha,\theta}(v_1, \ldots, v_n) dv_1 \cdots dv_n = E\left[\sum_{i_1,\ldots,i_n (\neq)} \delta_{(V_{i_1},\ldots,V_{i_n})}(dv_1 \cdots dv_n)\right]$$

is given by

$$q_{n,\alpha,\theta}(v_1, \ldots, v_n) = c_{n,\alpha,\theta} \prod_{i=1}^{n} v_i^{-(\alpha+1)} v_{n+1}^{\theta+\alpha n-1} \mathbf{1}_{\Delta_n}(v_1, \ldots, v_{n+1}),$$

where $c_{n,\alpha,\theta} = \prod_{i=1}^{n} B(1-\alpha, \theta+\alpha i)^{-1}$ **and**

$$\Delta_n = \left\{(v_1, \ldots, v_{n+1}) : v_1 \geq 0, \ldots, v_{n+1} \geq 0, v_1 + \cdots + v_{n+1} = 1\right\}.$$

(This generalizes Watterson's formula (1976) for $\alpha = 0$**.)**
<u>**Proof.**</u> **Use the aforementioned invariance of** $\mathrm{GEM}(\alpha, \theta)$**.**

Probability generating functional

Theorem 2 Let $(V_1, V_2, \ldots) \sim \mathsf{PD}(\alpha, \theta)$. Suppose that $g : (0, \infty) \to \mathbb{C}$ and $\lambda > 0$ are such that $\int_0^\infty |g(z) - 1| e^{-\lambda z} z^{-(\alpha+1)} dz < \infty$. Then

$$\lambda^\theta \int_0^\infty \frac{ds}{s^{1-\theta}} e^{-\lambda s} \left(E\left[\prod_{i=1}^{\infty} g(sV_i) \right] - 1 \right) = R_{\alpha,\theta} \left(\frac{1}{\lambda^\alpha} \int_0^\infty \frac{dz}{z^{1+\alpha}} e^{-\lambda z} (g(z) - 1) \right),$$

where

$$R_{\alpha,\theta}(u) := \begin{cases} \Gamma(\theta) \left(e^{\theta u} - 1 \right), & \alpha = 0, \theta > 0, \\ \Gamma(\theta + 1) \left\{ (1 - C_\alpha u)^{-\frac{\theta}{\alpha}} - 1 \right\} \theta^{-1}, & 0 < \alpha < 1, \theta \neq 0, \\ -\alpha^{-1} \log(1 - C_\alpha u), & 0 < \alpha < 1, \theta = 0 \end{cases}$$

and $C_\alpha = \alpha / \Gamma(1 - \alpha)$. (This refines a result of Pitman-Yor(1997).)

Proof. Put $g(v) =: 1 + \phi(v)$ and note that

$$\prod_{i=1}^{\infty} g(sV_i) = \prod_{i=1}^{\infty} (1 + \phi(sV_i)) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{i_1, \ldots, i_n (\neq)} \phi(sV_{i_1}) \cdots \phi(sV_{i_n}).$$

How to use Theorem 2 for $E\left[\prod_{i=1}^{\infty} g(sV_i)\right]$

- $g(s) = 1_{(0,1)}(s)$ $\Rightarrow E\left[\prod_{i=1}^{\infty} g(sV_i)\right] = P(sV_1 < 1) =: \rho_{\alpha,\theta}(s)$
  a generalization of Dickman's function
- $g(s) = e^{-s^p}$ $(p > 0)$ $\Rightarrow E\left[\prod_{i=1}^{\infty} g(sV_i)\right] = E\left[\exp\left(-s^p \sum_{i=1}^{\infty} V_i^p\right)\right]$
  Laplace transform of pth 'population moment'
- $g(s) = \int_{\mathbf{R}} e^{\pm\sqrt{-1}sx}\nu(dx)$ (ν: a given distribution)
  $\Rightarrow E\left[\prod_{i=1}^{\infty} g(sV_i)\right] = E\left[\exp\left(\pm\sqrt{-1}s \sum_{i=1}^{\infty} X_i V_i\right)\right]$
  the charactersitic function of a generalized Dirichlet mean
  $M := \sum_{i=1}^{\infty} X_i V_i$, where $\{X_i\}$: i.i.d. indep. of $(V_i)_{i=1}^{\infty}$, $X_i \sim \nu$.

For instance, we have

Theorem 3 $\rho_{\alpha,\theta}(s)$ sloves the equation

$$(s-1)^{\theta}\rho_{\alpha,\theta}(s-1) + \int_{s-1}^{s} (s-t)^{-\alpha} t^{\theta} d\rho_{\alpha,\theta}(t) = 0 \qquad (s > 1).$$

Known for $\theta = 0$ [Lamperti, 1961], for $\alpha = 0$ [Griffiths, 1979].

How to use Theorem 1 for $q_{n,\alpha,\theta}$

Accroding to the general theory, the density $P(V_m \in dv)/dv$ is given in terms of the correlation functions by

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(m-1)!k!} \int_{[v,1)^{m+k-1}} dy_1 \cdots dy_{m+k-1} q_{m+k,\alpha,\theta}(v, y_1, \ldots, y_{m+k-1}).$$

Moreover,

Theorem 4 **The joint density of $(V_1, \ldots, V_m)$ is**

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{[v_m,1)^k} dy_1 \cdots dy_k q_{m+k,\alpha,\theta}(v_1, \ldots, v_m, y_1, \ldots, y_k)$$

$$= q_{m,\alpha,\theta}(v_1, \ldots, v_m) \rho_{\alpha,\theta+\alpha m}\left(\frac{v_{m+1}}{v_m}\right)$$

for $(v_1, \ldots, v_{m+1}) \in \Delta_m$ such that $v_1 > \cdots > v_m$.
(This generalizes Watterson's result (1976) for $\alpha = 0$.)

Asymptotic results as $\theta \to \infty$

Fix $\alpha \in (0,1)$. For each $\theta > 1$, let $(V_i^{(\alpha,\theta)})_{i=1}^{\infty} \sim \mathrm{PD}(\alpha,\theta)$ and set $\beta_{\alpha,\theta} = \log\theta - (\alpha+1)\log\log\theta - \log\Gamma(1-\alpha)$.

Theorem 5 **As $\theta \to \infty$,**

$$\left(\theta V_i^{(\alpha,\theta)} - \beta_{\alpha,\theta}\right)_{i=1}^{\infty} \longrightarrow (Z_i^*)_{i=1}^{\infty} \sim \textbf{Poisson}(\mathbf{R}, e^{-z}dz)$$

in finite dimensional distributions sense.

Proof reduces to the corresponding asymptotics for $\rho_{\alpha,\theta}$.

Theorem 6 **As $\theta \to \infty$**

$$\left\{\sqrt{\theta}\left(\frac{\Gamma(1-\alpha)}{\Gamma(p-\alpha)}\theta^{p-1}\sum_{i=1}^{\infty}\left(V_i^{(\alpha,\theta)}\right)^p - 1\right) : p > \alpha\right\} \overset{\text{f.d.d.}}{\longrightarrow} \{W_p : p > \alpha\},$$

where $\{W_p\}$ is a centered Ganssian system with covariance $C(p,p') := \Gamma(1-\alpha)\Gamma(p+p'-\alpha)/(\Gamma(p-\alpha)\Gamma(p'-\alpha)) + \alpha - pp'$.

(Theorems 5 and 6 extend the known results for $\alpha = 0$ by Griffiths (1979) and Joyce et al. (2002), respectively.)

The generalized Dirichlet mean

Given a distribution ν, let $\{X_i\}$ be i.i.d. indep. of $(V_i^{(\alpha,\theta)})_{i=1}^{\infty}$ and $X_i \sim \nu$. Define $\mathcal{M}_{\alpha,\theta}\nu$ to be the law of $M := \sum_{i=1}^{\infty} X_i V_i^{(\alpha,\theta)}$ provided that it converges absolutely.
Taking $g(s) = \psi_\nu(\pm s) := \int_{\mathrm{R}} e^{\pm\sqrt{-1}sx}\nu(dx)$ in Theorem 2, we get
Proposition 7 **Assume that $\int_{\mathrm{R}} \log(1+|x|)\nu(dx) < \infty$ in case $\alpha = 0$ or $\int_{\mathrm{R}} |x|^\alpha \nu(dx) < \infty$ in case $0 < \alpha < 1$.**
Then $\mathcal{M}_{\alpha,\theta}\nu$ is well-defined and for all $\lambda > 0$

$$\lambda^\theta \int_0^\infty ds \frac{e^{-\lambda s}}{s^{1-\theta}} \left(\psi_{\mathcal{M}_{\alpha,\theta}\nu}(\pm s) - 1\right) = R_{\alpha,\theta}\left(\lambda^{-\alpha}\int_0^\infty dz \frac{e^{-\lambda z}}{z^{1+\alpha}}(\psi_\nu(\pm z) - 1)\right).$$

Moreover, for a.e. $t \in \mathrm{R}$,

$$\psi_{\mathcal{M}_{\alpha,\theta}\nu}(t) = 1 + \sum_{n=1}^{\infty} \frac{c_{n,\alpha,\theta}}{n!} \int_{\Delta_n} \prod_{i=1}^{n} \frac{\psi_\nu(tv_i) - 1}{v_i^{\alpha+1}} v_{n+1}^{\theta+\alpha n-1} dv_1 \cdots dv_n.$$