

随机向量的生成

王璐

在上一章我们介绍了生成一元随机变量的方法，本章将讨论如何生成 \mathbb{R}^d ($d > 1$) 上的随机向量。多元抽样的挑战在于如何给随机向量的元素之间赋予正确的相关结构。

对于一元随机变量，我们在上一章介绍了三种主要的抽样方法：CDF 逆变换、A-R 方法和混合抽样。它们都可以推广到多元情形，然而实践中除了几个成功的特例，使用这些方法进行多元抽样的效率通常很低。因此人们又提出了 Markov chain Monte Carlo, Sequential Monte Carlo 等方法，我们后面会介绍。

本章我们将重点关注一些常用的多元分布，比如多元正态分布、多元 t 分布、Dirichlet 分布以及多项分布 (multinomial distribution) 等，对这些多元分布进行抽样已经有非常高效的方法。

除了介绍上述几种多元分布的抽样方法，我们还将介绍一种更一般的多元抽样方法 — copula-marginal 方法。它可以看作一元的 QQ 变换推广到多元的方法，其基本想法是将一种相关结构已知的多元分布通过边际分布变换得到另一多元分布。当然这一过程是很复杂的，因此人们也将这一方法视为第四种主要抽样方法。

最后我们还会介绍一些随机矩阵的抽样方法。这些随机矩阵可以看作若干有相关结构的随机向量的集合，比如在球面上随机分布的点，Wishart 矩阵，随机图等。

本章我们假设维度 $d < \infty$ ，下一章我们将讨论 $d = \infty$ 的情况，即随机过程的产生。

1 一元抽样方法的推广

本节我们将讨论如何将三种主要的一元抽样方法 — CDF 逆变换，A-R 方法，混合抽样 — 推广到多元抽样。

1.1 CDF 逆变换

直接抽取随机向量 $\mathbf{X} \in \mathbb{R}^d$ 的方法是依次抽取它的各元素 X_1, X_2, \dots, X_d 。首先从 X_1 的边际分布抽样，然后给定已有元素的样本，从条件分布中抽取下一个元素的样本，原理是

$$f_{\mathbf{X}}(\mathbf{x}) = f_1(x_1)f_{2|1}(x_2 | x_1)f_{3|1:2}(x_3 | x_{1:2}) \cdots f_{d|1:(d-1)}(x_d | x_{1:(d-1)}). \quad (1)$$

根据序列生成形式(1)，我们可以构造如下的 CDF 逆变换法在 d 维空间抽样，称为 sequential inversion. 首先抽取 d 个 $U(0, 1)$ 的样本 $U_j \stackrel{iid}{\sim} U(0, 1)$, $j = 1, 2, \dots, d$. 然后依次令

$$X_1 = F_1^{-1}(U_1)$$

$$X_j = F_{j|1:(j-1)}^{-1}(U_j | X_{1:(j-1)}), \quad j = 2, \dots, d.$$

就得到随机向量 \mathbf{X} 的一个样本。可以看到，使用 sequential inversion 需要知道 X_1 边际分布的 CDF 以及其它序列条件分布的 CDF. 我们来看一个具体例子。

• **Example.** 随机向量 $\mathbf{X} = (X_1, X_2)$ 的 PDF 为

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & (x_1, x_2) \in [0, 1]^2 \\ 0, & \text{else.} \end{cases}$$

下面用 sequential inversion 对 \mathbf{X} 进行抽样：

– X_1 的边际 PDF 为 $f_1(x_1) = \int_0^1 f(x_1, x_2) dx_2 = x_1 + 1/2$. 因此 X_1 的边际 CDF 为 $F_1(x) = \int_0^x f_1(t) dt = (x^2 + x)/2$, $0 \leq x \leq 1$. 利用二次方程求根公式，可得 F_1 的逆函数：

$$X_1 = F_1^{-1}(U_1) = \sqrt{2U_1 + 1/4} - 1/2, \quad U_1 \sim U(0, 1).$$

– 给定 $X_1 = x_1$, X_2 的条件分布的 PDF 为

$$f_{2|1}(x_2 | x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{x_1 + x_2}{x_1 + 1/2}, \quad 0 \leq x_2 \leq 1.$$

X_2 的条件分布的 CDF 为

$$F_{2|1}(x | x_1) = \int_0^x f_{2|1}(t | x_1) dt = \frac{x_1 x + x^2/2}{x_1 + 1/2}.$$

再次利用二次方程求根公式，可得 $F_{2|1}$ 的逆变换为：

$$X_2 = F_{2|1}^{-1}(U_2 | X_1) = \sqrt{X_1^2 + (2X_1 + 1)U_2} - X_1, \quad U_2 \sim U(0, 1).$$

Remarks

1. 使用 sequential inversion 进行多元抽样在实践中经常面临的问题是：序列条件分布 $F_{j|1:(j-1)}(x_j | x_{1:(j-1)})$ 的逆函数在高维情况下很难计算。而且如果每个条件分布的逆函数都需要重新计算，不能利用前面的结果，使用 sequential inversion 抽样会很慢。

2. 也有一些分布使用 sequential inversion 抽样很容易，比如后面会介绍的多项分布。

1.2 Acceptance-rejection (A-R)

A-R 方法也很容易推广到多元。如果想从 \mathbb{R}^d 上的分布 $f(\mathbf{x})$ (PDF) 中抽样, 可以先从另一个容易抽样的分布 $g(\mathbf{x})$ (PDF) 中抽样, 只要保证存在常数 c 使得 $f(\mathbf{x}) \leq cg(\mathbf{x})$. 容易证明在多元情形下, A-R 中来自 g 的样本总体被接受的概率也是 $1/c$. 即平均从 g 中抽取 c 个样本, 才有一个被接受作为 f 的样本。

A-R 的几何解释在多元情形下依然成立。令

$$S_c(f) = \left\{ (\mathbf{x}, z) \mid 0 \leq z \leq cf(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d \right\}$$

表示一个 $(d+1)$ 维的闭集。如果 $(\mathbf{X}, Z) \sim U(S_c(f))$, 则 $\mathbf{X} \sim f$. 反过来, 如果随机向量 $\mathbf{X} \sim f$ 且 $Z \mid \mathbf{X} = \mathbf{x} \sim U(0, cf(\mathbf{x}))$, 则 $(\mathbf{X}, Z) \sim U(S_c(f))$.

A-R 的几何解释保证了我们可以使用 f 和 g 未归一化的形式 — \tilde{f} 和 \tilde{g} — 计算来自 g 的样本被接受的概率:

$$\mathbf{Y} \sim g, A(\mathbf{Y}) = \frac{\tilde{f}(\mathbf{Y})}{\tilde{c}\tilde{g}(\mathbf{Y})}$$

只要保证 $\tilde{f}(\mathbf{y}) \leq \tilde{c}\tilde{g}(\mathbf{y}), \forall \mathbf{y}$.

- **Example.** 目标分布 f 是单位球体 $B_d = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| \leq 1\}$ 内的均匀分布, 令 g 表示 $U[-1, 1]^d$ 的 PDF. 它们的 unnormalized PDF 为 $\tilde{f}(\mathbf{x}) = \mathbf{1}(\mathbf{x} \in B_d)$ 和 $\tilde{g}(\mathbf{x}) = \mathbf{1}(\mathbf{x} \in [-1, 1]^d)$, 因此在 A-R 中选取 $\tilde{c} = 1$ 即可。此时抽取 $\mathbf{Y} \sim g$ 后只保留 $\|\mathbf{Y}\| \leq 1$ 的样本, 则来自 g 的样本总体被接受的概率为

$$\frac{\text{vol}(B_d)}{2^d} = \frac{\pi^{d/2}}{2^d \Gamma(1 + d/2)}.$$

- $d = 2$ 时, 上述接受概率为 $\pi/4 \approx 0.785$, 比较高。
- $d = 9$ 时, 上述接受概率 $< 1\%$; $d = 23$ 时, 接受概率 $< 10^{-9}$.

- **Example.** 假设 f 和 g 都可写为 d 个一元 PDF 的乘积 (各分量独立), 且存在 $c_j = \sup_{\mathbf{x}} f_j(\mathbf{x})/g_j(\mathbf{x}), j = 1, 2, \dots, d$. 则在 A-R 中可选取常数 c 为

$$c = \sup_{\mathbf{x}} f(\mathbf{x})/g(\mathbf{x}) = \prod_{j=1}^d c_j$$

如果每个 $c_j > 1 + \epsilon$, 则 c 将随着 d 指数增长。

Remark

- 通过上述例子, 可以看到 A-R 方法在多元抽样中经常面临的问题是:
 - 在高维情形下一般很难找到较小的 c , 抽样效率很低。
 - 计算 $c = \sup_{\mathbf{x}} f(\mathbf{x})/g(\mathbf{x})$ 很复杂, 一般需要解一个 d 维优化。

1.3 混合抽样

混合抽样 (mixture sampling) 很容易推广到高维, 有时也能使多元抽样变得简单。如果多元分布的 PDF 可以写为如下的连续混合形式

$$f(\mathbf{x}) = \int f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y})g(\mathbf{y})d\mathbf{y},$$

则可先抽取 $\mathbf{Y} \sim g$, 给定 \mathbf{Y} 再从条件分布抽样 $\mathbf{X} | \mathbf{Y} \sim f_{\mathbf{X}|\mathbf{Y}}$, 即得 $\mathbf{X} \sim f$ 的样本。

如果 $f(\mathbf{x})$ 可写为如下的离散混合形式

$$f(\mathbf{x}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x})$$

其中 $\pi_k \geq 0$ 且 $\sum_{k=1}^K \pi_k = 1$; 每个 f_k 都是 \mathbb{R}^d 上的一个 PDF. 则可如下对 f 抽样: 先对一个离散的随机变量 Z 抽样, $P(Z = k) = \pi_k, k = 1, 2, \dots, K$; 给定 $Z = k$, 再从 f_k 抽样 $\mathbf{X} | Z = k \sim f_k$.

2 多元正态分布 (Multivariate normal)

多元正态分布是最重要的多元分布之一。 \mathbb{R}^d 上的多元正态分布由一个期望向量 $\boldsymbol{\mu} \in \mathbb{R}^d$ 和一个半正定的协方差矩阵 $\Sigma \in \mathbb{R}^{d \times d}$ 决定, 记为 $N_d(\boldsymbol{\mu}, \Sigma)$. 如果 Σ 可逆, $N_d(\boldsymbol{\mu}, \Sigma)$ 的 PDF 为

$$\phi(\mathbf{x} | \boldsymbol{\mu}, \Sigma) = \frac{\exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}}{(2\pi)^{d/2} |\Sigma|^{1/2}}, \mathbf{x} \in \mathbb{R}^d$$

如果 Σ 不可逆, 说明有些成分是多余的, 即存在 $k \in \{1, \dots, d\}$ 满足

$$P(X_k = \alpha_0 + \sum_{j \neq k} \alpha_j X_j) = 1.$$

下面我们只讨论没有多余成分的正态分布, 即 Σ 是可逆的情形。

多元正态分布有很多重要性质 (后面会经常用到):

1. 如果 $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, 则 $A\mathbf{X} + \mathbf{b} \sim N_d(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^\top)$.
2. 将 \mathbf{X} 分为不相交的两个子向量 $\mathbf{X}_1 = (X_1, \dots, X_r)$ 和 $\mathbf{X}_2 = (X_{r+1}, \dots, X_d)$, 对参数也做相应地划分

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

则它们各自的边际分布为 $\mathbf{X}_j \sim N(\boldsymbol{\mu}_j, \Sigma_{jj}), j = 1, 2$.

3. 上述 \mathbf{X}_1 和 \mathbf{X}_2 独立当且仅当 Σ_{12} 是零矩阵。

4. 给定 $\mathbf{X}_2 = \mathbf{x}_2$, \mathbf{X}_1 的条件分布为

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim N_d(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

从 $N_d(\mathbf{0}, I_d)$ 中抽样很容易, 因为此时各分量的相关性都为 0, 在多元正态分布中, 这意味着各分量都是独立的。因此可以使用 Box-Muller 或 CDF 逆变换独立地从 $N(0, 1)$ 中抽样, $Z_j \stackrel{iid}{\sim} N(0, 1)$, $j = 1, \dots, d$, 则 $\mathbf{Z} = (Z_1, \dots, Z_d)^\top \sim N_d(\mathbf{0}, I_d)$.

对一般的多元正态分布 $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ 抽样, 只需找到矩阵 C 使得 $\Sigma = CC^\top$, 然后对 \mathbf{Z} 做线性变换即可:

$$\mathbf{X} = \boldsymbol{\mu} + C\mathbf{Z}, \quad \mathbf{Z} \sim N_d(\mathbf{0}, I_d).$$

上述矩阵 C 总可以通过特征值分解获得。由于 Σ 是对称的半正定矩阵, 因此存在特征值分解

$$\Sigma = P\Lambda P^\top$$

其中 Λ 是对角阵且对角线元素非负 $\Lambda_{ii} \geq 0$. 因此可令 $C = P\Lambda^{1/2}P^\top$. 矩阵 C 的选择并不唯一, 对于任意正交矩阵 Q , 令 $\tilde{C} = CQ$, 则 $\tilde{C}\tilde{C}^\top = CQQ^\top C^\top = CC^\top = \Sigma$.

由于 Σ 是半正定的, 人们也经常使用 Cholesky 分解 $\Sigma = LL^\top$, 然后令 $C = L$, 其中 L 是下三角矩阵。当 Σ 正定时, Cholesky 分解是唯一的, 此时 L 的对角线元素全部为正。

特征值分解和 Cholesky 分解的计算量都是 $O(d^3)$.

3 多元 t 分布

多元 t 分布有三部分参数, center $\boldsymbol{\mu}$, scale matrix Σ 和自由度 ν , 记为 $t_d(\boldsymbol{\mu}, \Sigma, \nu)$. 当 $\nu = 1$ 时, 多元 t 分布也称多元 Cauchy 分布; 当 $\nu \rightarrow \infty$ 时, $t_d(\boldsymbol{\mu}, \Sigma, \nu)$ 收敛到 $N_d(\boldsymbol{\mu}, \Sigma)$.

\mathbb{R}^d 上的 $t_d(\boldsymbol{\mu}, \Sigma, \nu)$ 的 PDF 为

$$f(\mathbf{x} | \boldsymbol{\mu}, \Sigma, \nu) = C_{\boldsymbol{\mu}, \Sigma, \nu} (1 + (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))^{-(\nu+d)/2}$$

其中归一化常数为

$$C_{\boldsymbol{\mu}, \Sigma, \nu} = \frac{\Gamma((\nu+d)/2)}{|\Sigma|^{1/2} (\nu\pi)^{d/2} \Gamma(\nu/2)}.$$

对于标准的多元 t 分布, $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I$, 此时 $f(\mathbf{x}) \propto (1 + \|\mathbf{x}\|^2)^{-(\nu+d)/2}$. 但是 $\Sigma = I_d$ 的多元 t 分布的各分量并不独立。

$t_d(\boldsymbol{\mu}, \Sigma, \nu)$ 的各分量的边际分布为

$$\frac{X_j - \mu_j}{\sqrt{\Sigma_{jj}}} \sim t_{(\nu)}.$$

与多元正态分布相似, 多元 t 分布 $t_d(\boldsymbol{\mu}, \Sigma, \nu)$ 的 PDF 的形状是以 $\boldsymbol{\mu}$ 为中心的一系列椭圆等高线, 但多元 t 分布依然比多元正态分布的尾厚。多元 t 分布可由如下变换生成:

$$\mathbf{X} = \boldsymbol{\mu} + \frac{\Sigma^{1/2} \mathbf{Z}}{\sqrt{W/\nu}}, \quad \mathbf{Z} \sim N_d(\mathbf{0}, I_d), W \sim \chi^2_{(\nu)}$$

其中 \mathbf{Z} 和 W 独立, $\Sigma^{1/2}$ 是任何满足 $CC^\top = \Sigma$ 的矩阵 C 。

4 多项分布 (Multinomial)

如果向 d 个格子独立地抛 m 个球, 每个球落入格子 j 的概率为 p_j , $j = 1, \dots, d$. 则落入每个格子 j 的球数 X_j 组成的向量 $\mathbf{X} = (X_1, \dots, X_d)$ 服从多项分布 $\text{Mult}(m, p_1, \dots, p_d)$. 它的 PMF 为

$$P(X_1 = x_1, \dots, X_d = x_d) = \frac{m!}{x_1! x_2! \dots x_d!} \prod_{j=1}^d p_j^{x_j}$$

其中 x_j 为非负整数且满足 $\sum_{j=1}^d x_j = m$, 概率 $p_j \geq 0$ 且 $\sum_{j=1}^d p_j = 1$. 因此参数向量 $\mathbf{p} = (p_1, \dots, p_d)$ 可取值的集合为

$$\Delta^{d-1} = \left\{ (p_1, \dots, p_d) \mid p_j \geq 0, \sum_{j=1}^d p_j = 1 \right\}$$

Δ^{d-1} 被称为 \mathbb{R}^d 上的 unit simplex. Δ 的上标 $d-1$ 表示该集合的真实维度是 $d-1$.

对多项分布抽样可以按如下序列条件分布的形式依次对每个分量抽样

$$P(X_1, \dots, X_d) = P(X_1)P(X_2 \mid X_1) \cdots P(X_j \mid X_1, \dots, X_{j-1}) \cdots P(X_d \mid X_1, \dots, X_{d-1})$$

其中 X_1 的边际分布是一个二项分布 $X_1 \sim \text{Bin}(m, p_1)$; 给定 $\{X_1, \dots, X_{j-1}\}$, X_j 的条件分布也是一个二项分布: 此时可能落入格子 j 的球数变为 $m - \sum_{s=1}^{j-1} X_s$, 且这些球只能落入格子 j, \dots, d , 因此每个球落入格子 j 的概率增大为 $p_j / \sum_{k=j}^d p_k$, 所以

$$X_j \mid X_1, \dots, X_{j-1} \sim \text{Bin} \left(m - \sum_{s=1}^{j-1} X_s, p_j / \sum_{k=j}^d p_k \right).$$

上述抽样方法可以用算法1实现。

Algorithm 1 Sample $\mathbf{X} \sim \text{Mult}(m, p_1, \dots, p_d)$

 Input $m \in \mathbb{N}$, $d \in \mathbb{N}$, $\mathbf{p} = (p_1, \dots, p_d) \in \Delta^{d-1}$.

 Let $n = m$ and $S = 1$.

for $j = 1$ to d **do**
 $X_j = \text{Bin}(n, p_j/S)$
 $n = n - X_j$ ▷ 如果在某步迭代中发现 $n = 0$, 则可直接将后面的分量取为 0.
 $S = S - p_j$
return $\mathbf{X} = (X_1, \dots, X_d)$

5 Dirichlet 分布

有时要抽样的随机向量可能是一组随机概率，比如从多项分布的参数空间抽样，此时抽样的样本空间是一个 unit simplex

$$\Delta^{d-1} = \left\{ (x_1, \dots, x_d) \mid x_j \geq 0, \sum_{j=1}^d x_j = 1 \right\}$$

Dirichlet 分布是定义在 unit simplex Δ^{d-1} ($d \geq 2$) 上最简单的分布之一，它有 d 个参数： $\alpha_j > 0$, $j = 1, \dots, d$, 记为 $\text{Dir}(\alpha_1, \dots, \alpha_d)$. 它的 PDF 为

$$f(\mathbf{x}) = \frac{1}{D(\boldsymbol{\alpha})} \prod_{j=1}^d x_j^{\alpha_j-1}, \quad \mathbf{x} \in \Delta^{d-1}$$

其中归一化常数 $D(\boldsymbol{\alpha}) = \prod_{j=1}^d \Gamma(\alpha_j) / \Gamma(\sum_{j=1}^d \alpha_j)$. $\mathbf{X} \sim \text{Dir}(\boldsymbol{\alpha})$ 的期望为

$$E(X_j) = \frac{\alpha_j}{\sum_{k=1}^d \alpha_k}, \quad j = 1, \dots, d.$$

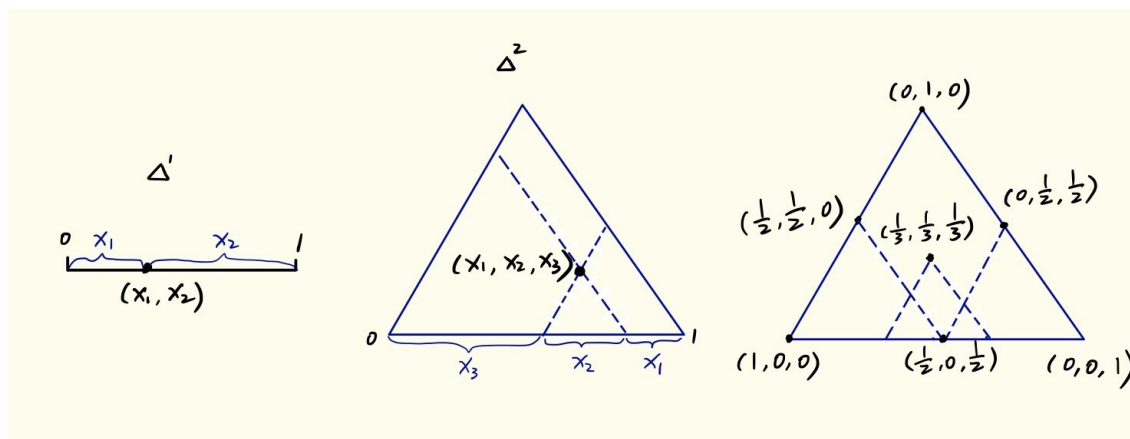
Dirichlet 分布有两个特例值得说明：

1. $d = 2$ 时的 Dirichlet 分布 $\text{Dir}(\alpha_1, \alpha_2)$ 等价于 $\text{Beta}(\alpha_1, \alpha_2)$ 分布，即

$$(X_1, X_2) \sim \text{Dir}(\alpha_1, \alpha_2) \Leftrightarrow X_1 \sim \text{Beta}(\alpha_1, \alpha_2), X_2 = 1 - X_1$$

且此时 $X_2 \sim \text{Beta}(\alpha_2, \alpha_1)$.

2. $\alpha_j \equiv 1$, $j = 1, \dots, d$ 对应的 Dirichlet 分布是 Δ^{d-1} 上的均匀分布 $U(\Delta^{d-1})$.

Figure 1: Δ^1 和 Δ^2 空间

$d = 2$ 时的样本空间 Δ^1 对应一个长度为 1 的线段, $d = 3$ 时的样本空间 Δ^2 可以用一个等边三角形表示, 如图1所示。

图2展示了 6 组不同参数向量 $\alpha \in \mathbb{R}_+^3$ 对应的 $\text{Dir}(\alpha)$ 的样本。可以看到, 样本倾向于分布在最大的 α_j 对应的角附近。比较 $\text{Dir}(1, 1, 1)$, $\text{Dir}(7, 7, 7)$ 和 $\text{Dir}(0.2, 0.2, 0.2)$ 的样本分布, 虽然这三个分布的期望相同, 但样本的表现却很不同: $\alpha_j \equiv 1$ 对应 Δ^2 上的均匀分布; 较大的 α_j 's 倾向于让样本更靠近中心, 即分布的期望; 较小的 α_j 's 倾向于让样本更靠近边界, 边界上的点会有某个分量为 0。

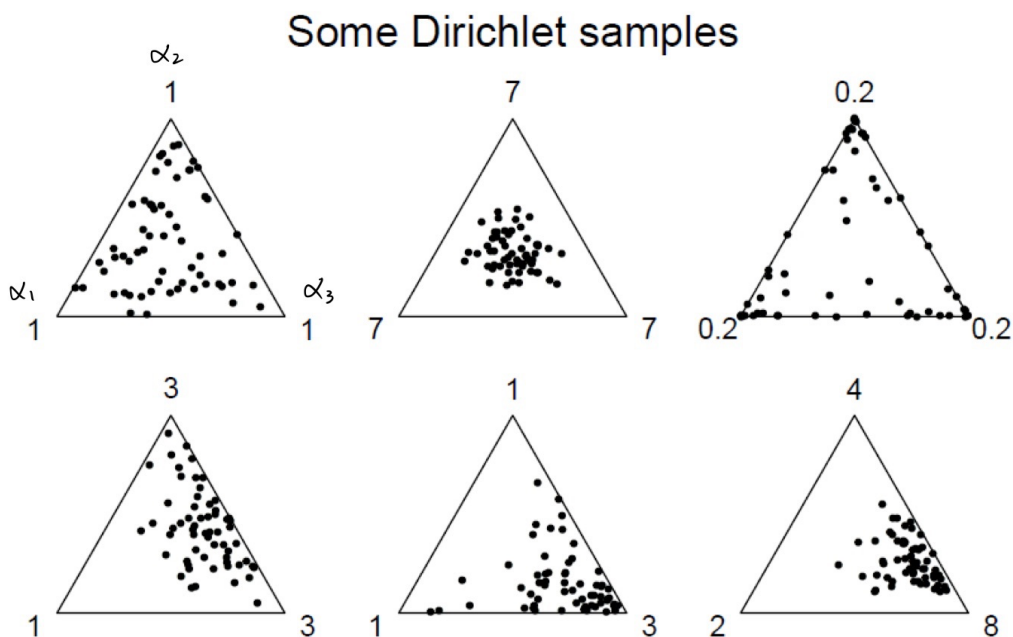


Figure 2: 6 组不同 $\alpha \in \mathbb{R}_+^3$ 对应的 $\text{Dir}(\alpha)$ 样本 (每组 60 个)。 $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ 的取值标在三角形的各角上。 Picture source: Art B. Owen.

我们在上一章介绍了用 Gamma 分布生成 Beta 分布的方法。类似地, Dirichlet 分布也可以由 Gamma 分布生成, 方法如下:

$$\begin{aligned} Y_j &\stackrel{iid}{\sim} \text{Gam}(\alpha_j), j = 1, \dots, d, \quad \text{then} \\ X_j &= \frac{Y_j}{\sum_{k=1}^d Y_k}, j = 1, \dots, d. \end{aligned} \quad (2)$$

则 $\mathbf{X} \sim \text{Dir}(\boldsymbol{\alpha})$, 其中 $\text{Gam}(\alpha_j) = \text{Gam}(\alpha_j, 1)$. 由(2)可得 $\text{Dir}(\boldsymbol{\alpha})$ 的边缘分布为

$$X_j \sim \text{Beta}(\alpha_j, \sum_{k \neq j} \alpha_k), j = 1, \dots, d. \quad (3)$$

因为根据 Gamma 分布的性质, $Y_{-j} = \sum_{k \neq j} Y_k \sim \text{Gam}(\sum_{k \neq j} \alpha_k)$ 且与 Y_j 独立, 而 $X_j = Y_j / (Y_j + Y_{-j})$, 所以得到(3). 因此 Dirichlet 分布也可以看作多元 Beta 分布。

- 对于 $\alpha_j \equiv 1, j = 1, \dots, d$ 的 Dirichlet 分布, 即 $\mathbf{U}(\Delta^{d-1})$, 此时 $\text{Gam}(1)$ 即为 $\text{Exp}(1)$ 分布, 而 $\text{Exp}(1)$ 的样本可由变换 $Y_j = -\log(U_j), U_j \sim \mathbf{U}(0, 1)$ 得到, 则可令

$$X_j = \frac{\log(U_j)}{\sum_{k=1}^d \log(U_k)}, U_j \stackrel{iid}{\sim} \mathbf{U}(0, 1).$$

- $\mathbf{U}(\Delta^{d-1})$ 还可以使用 **uniform spacings** 方法抽样。令

$$U_j \stackrel{iid}{\sim} \mathbf{U}(0, 1), j = 1, \dots, d-1$$

它们对应的 order statistics 为 $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(d-1)}$. 再扩展两个点 $U_{(0)} = 0, U_{(d)} = 1$, 然后令

$$X_j = U_{(j)} - U_{(j-1)}, j = 1, \dots, d$$

则 $\mathbf{X} \sim \mathbf{U}(\Delta^{d-1})$. 该方法只需产生 $d-1$ 个随机变量且避免了对数运算, 但是排序的计算量为 $O(d \log(d))$, 因此对很大的 d , uniform spacings 可能比从指数分布抽样慢。

Remarks

1. Dirichlet 分布不是很灵活的分布, 它只有 d 个参数, 而期望 $E(\mathbf{X}) = \boldsymbol{\alpha} / \sum_{j=1}^d \alpha_j$ 用掉了 $d-1$ 个参数, 剩下的归一化参数 $\sum_{j=1}^d \alpha_j$ 描述 \mathbf{X} 距 $E(\mathbf{X})$ 的远近。因此没有足够的参数让 \mathbf{X} 各分量的方差自由变化, 更不用说它们之间的 $d(d-1)/2$ 对相关关系。
2. Dirichlet 分布的各分量几乎是独立的, 由于和为 1 的限制, 各分量间有很小的负相关。因此不能用 Dirichlet 分布产生 Δ^{d-1} 上分量间有正相关的样本, 或是分量间存在很大负相关的样本。