

1.2 HW 2

作业 2 链接

练习 1.1 假设 $\mathbf{x} = (x_1, \dots, x_n)^\top$ 服从球对称分布，证明其部分 $(x_1, \dots, x_k)^\top, 1 \leq k < n$ 也服从球对称分布。

证明 记 $\mathbf{x}_{i:j} = (x_i, \dots, x_j)^\top, 1 \leq i \leq j \leq n$. 考虑正交矩阵 $\tilde{O} = \text{diag}(O, I_{n-k}) \in \mathbb{R}^{n \times n}$, 其中 $O \in \mathbb{R}^{k \times k}$. 由 $\mathbf{x}$ 是球对称分布可知, $\mathbf{x} \stackrel{d}{=} \tilde{O}\mathbf{x} = (O\mathbf{x}_{1:k}, \mathbf{x}_{k+1:n})^\top$. 故对于任意正交阵 $O \in \mathbb{R}^{k \times k}$, $\mathbf{x}_{1:k} \stackrel{d}{=} O\mathbf{x}_{1:k}$, 因此 $\mathbf{x}_{1:k}$ 也服从球对称分布.

练习 1.2 假设随机向量 $(x, y)^\top \in B^2$, 做极坐标变换

$$x = r \cos(\theta), \quad y = r \sin(\theta), \quad r > 0, \quad 0 \leq \theta < 2\pi,$$

假设 $\theta \sim U(0, 2\pi), r \sim \text{Beta}(1, \beta) (\beta > 0)$, 试求 (x, y) 的联合概率密度; 当 $\beta \rightarrow \infty$ 时, 该分布有什么特征?

证明 Jacobi 矩阵行列式为:

$$|J| = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r.$$

所以 (x, y) 的联合密度函数为:

$$f(x, y) = p(r, \theta) |J|^{-1} = p_\theta(\theta) \cdot p_r(r) \cdot |J|^{-1} = \frac{1}{2\pi} \cdot \frac{1}{B(1, \beta)} (1-r)^{\beta-1} \cdot \frac{1}{r} = \frac{\beta(1-r)^{\beta-1}}{2\pi r} = \frac{\beta(1-\sqrt{x^2+y^2})^{\beta-1}}{2\pi \sqrt{x^2+y^2}}.$$

当 $\beta \rightarrow \infty$ 时, 容易观察到分布集中在 $(0, 0)$ 处。

练习 1.3 假设三元随机向量 $(x, y, z)^\top$ 服从球对称分布, 其概率密度函数 $f(x, y, z) = h(\sqrt{x^2 + y^2 + z^2})$, 令 $(x, y, z)^\top$ 的球坐标变换

$$x = r \cos(\theta), \quad y = r \sin(\theta) \cos(\psi), \quad z = r \sin(\theta) \sin(\psi), \quad r > 0, \quad 0 \leq \theta < \pi, \quad 0 \leq \psi < 2\pi.$$

(a) 证明 $\cos(\theta) \sim U(-1, 1), \psi \sim U(0, 2\pi)$ 。

(b) 如果已知 $r \sim U(0, 1)$, 试求 $(x, y, z)^\top$ 的概率密度。

证明

(a) Jacobi 矩阵行列式为:

$$|J| = \begin{vmatrix} \sin \psi \cos \theta & r \cos \psi \cos \theta & -r \sin \psi \sin \theta \\ \sin \psi \sin \theta & r \cos \psi \sin \theta & r \sin \psi \cos \theta \\ \cos \psi & -r \sin \psi & 0 \end{vmatrix} = r^2 \sin \theta$$

所以 (r, θ, ψ) 的联合密度函数为:

$$p(r, \theta, \psi) = f(x, y, z) |J| = r^2 \sin \theta \cdot h(r) = \left(\frac{1}{2\pi}\right) \left(\frac{1}{2} \sin \theta\right) (4\pi r^2 h(r))$$

可知 r, θ, ψ 相互独立, 密度函数为:

$$p(r) = 4\pi r^2 h(r), \quad p(\psi) = \frac{1}{2\pi}, \quad p(\theta) = \frac{1}{2} \sin \theta$$

所以 $\cos(\theta) \sim U(-1, 1), \psi \sim U(0, 2\pi)$

(b) 由于 $r \sim U(0, 1)$, 并且 $p(r) = 4\pi r^2 h(r)$, 故 $(x, y, z)^\top$ 的密度函数为:

$$f(x, y, z) = \frac{1}{4\pi} \cdot \frac{1}{x^2 + y^2 + z^2}$$

练习 1.4 假设随机变量 $x \sim \text{Beta}(1/2, n/2)$, 令 $r = \sqrt{(1-x)/x}$. 假设 $\mathbf{u} \sim U(S^{n-1})$ 且与 r 独立, 试验证 $\mathbf{x} = r\mathbf{u}$ 的密度函数为

$$p(\mathbf{x}) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \cdot \frac{1}{(1 + \|\mathbf{x}\|^2)^{(n+1)/2}}, \quad \mathbf{x} \in \mathbb{R}^n.$$

这是多元 Cauchy 分布。

证明 由 $x = \frac{1}{1+r^2}$, 可知

$$|J| = \left| \frac{\partial \mathbf{x}}{\partial r} \right| = \left| \frac{-2r}{(1+r^2)^2} \right| = \frac{2r}{(1+r^2)^2}$$

故 r 的密度函数为:

$$p(r) = \left(\frac{1}{B\left(\frac{1}{2}, \frac{n}{2}\right)} x^{-1/2} (1-x)^{n/2-1} \right) \bigg|_{x=(1+r^2)^{-1}} |J| = \frac{\Gamma(n+1/2)}{\Gamma(1/2)\Gamma(n/2)} \cdot \frac{r^{n-1}}{(1+r^2)^{(n+1)/2}}.$$

由于 $\mathbf{u} \sim U(S^{n-1})$ 且根据第二讲定理 1, 可知:

$$p(\mathbf{x}) = \frac{\Gamma(n/2) \cdot p(\|\mathbf{x}\|)}{\pi^{n/2} \cdot \|\mathbf{x}\|^{n-1}} = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \cdot \frac{1}{(1+\|\mathbf{x}\|^2)^{(n+1)/2}}.$$

 **练习 1.5** 假设 k 元随机向量 $\mathbf{x}$ 服从如下分布

$$f(\mathbf{x}) = \frac{\Gamma(n/2)}{\pi^{k/2}\Gamma((n-k)/2)} (1 - \|\mathbf{x}\|^2)^{(n-k-2)/2}, \quad \mathbf{x} \in B^k, \quad k < n,$$

这是一个单位球 B^k 内的球对称分布 (实际上是球面均匀分布 $U(S^{n-1})$ 的 k 元边际分布, 参见第二讲定理 2), 试应用第二讲定理 1 (模长的概率密度公式) 证明 $\|\mathbf{x}\|^2 \sim \text{Beta}(k/2, (n-k)/2)$ 。

证明 由题设密度函数

$$f(\mathbf{x}) = \frac{\Gamma(n/2)}{\pi^{k/2}\Gamma((n-k)/2)} (1 - \|\mathbf{x}\|^2)^{(n-k-2)/2}$$

可知 $\mathbf{x}$ 是球对称分布, 满足定理 1 的条件, 即 $f(\mathbf{x}) = h(\|\mathbf{x}\|)$ 。因此模长 $r = \|\mathbf{x}\|$ 的密度为

$$\begin{aligned} p(r) &= |S^{k-1}| \cdot h(r) \cdot r^{k-1} = \frac{2\pi^{k/2}}{\Gamma(k/2)} \cdot \frac{\Gamma(n/2)}{\pi^{k/2}\Gamma((n-k)/2)} (1-r^2)^{(n-k-2)/2} \cdot r^{k-1} \\ &= \frac{2\Gamma(n/2)}{\Gamma(k/2)\Gamma((n-k)/2)} r^{k-1} (1-r^2)^{(n-k-2)/2}, \quad 0 < r < 1 \end{aligned}$$

令 $s = r^2$, 则 $\|\mathbf{x}\|^2 = s$ 的密度为

$$p(s) = p(r) \cdot \left| \frac{dr}{ds} \right| = \frac{\Gamma(n/2)}{\Gamma(k/2)\Gamma((n-k)/2)} s^{k/2-1} (1-s)^{(n-k)/2-1}$$

这正是 Beta 分布的密度形式。因此有

$$\|\mathbf{x}\|^2 \sim \text{Beta}\left(\frac{k}{2}, \frac{n-k}{2}\right)$$

 **练习 1.6** 设 $\mathbf{u} = (u_1, \dots, u_n)^\top \sim U(S^{n-1})$, 利用事实: $\mathbf{u} \stackrel{d}{=} \mathbf{z}/\|\mathbf{z}\|$, $\mathbf{z} \sim N_n(0, I_n)$, 证明 $u_1^2 + \dots + u_k^2 \sim \text{Beta}(k/2, (n-k)/2)$, ($k < n$)。

证明 令

$$Z_1 = z_1^2 + \dots + z_k^2 \sim \chi_k^2, \quad Z_2 = z_{k+1}^2 + \dots + z_n^2 \sim \chi_{n-k}^2$$

由于 $\mathbf{z}$ 的分量独立, 故 $Z_1 \perp Z_2$ 且 $\|\mathbf{z}\|^2 = Z_1 + Z_2 \sim \chi_n^2$ 。

从而

$$u_1^2 + \dots + u_k^2 \stackrel{d}{=} \frac{z_1^2 + \dots + z_k^2}{\|\mathbf{z}\|^2} = \frac{Z_1}{Z_1 + Z_2} \sim \text{Beta}\left(\frac{k}{2}, \frac{n-k}{2}\right)$$

 **练习 1.7**

(a) 假设 $x, y \text{ iid} \sim N(0, 1)$, 证明或说明 $x/y \sim t_1$ (即 Cauchy 分布)。

(b) 假设 $\mathbf{x} = (x_1, \dots, x_n)^\top \sim U(S^{n-1})$ (单位球面均匀分布), 证明 $x_1/x_2 \sim t_1$ 。

(c) 假设 $\mathbf{x} = (x_1, \dots, x_n)^\top \sim U(B^n)$ (单位球内均匀分布), 证明 $x_1/x_2 \sim t_1$ 。

证明 (a) 由于 $x \perp y$, 所以

$$\frac{x}{y} \stackrel{d}{=} \frac{x}{|y|} \sim t_1$$

即 Cauchy 分布。

(b) 若 $\mathbf{x} \sim U(S^{n-1})$, 则根据球面对称分布的性质, 有

$$\mathbf{x} \stackrel{d}{=} \frac{\mathbf{z}}{\|\mathbf{z}\|}, \quad \mathbf{z} \sim N_n(0, I_n)$$

从而

$$\frac{x_1}{x_2} \stackrel{d}{=} \frac{z_1}{z_2} \sim t_1$$

(c) 若 $\mathbf{x} \sim U(B^n)$, 则可表示为 $\mathbf{x} = r\mathbf{u}$, 其中 $r \in (0, 1)$, $\mathbf{u} \sim U(S^{n-1})$, 从而

$$\frac{x_1}{x_2} = \frac{ru_1}{ru_2} = \frac{u_1}{u_2}$$

由 (b) 已知 $\frac{u_1}{u_2} \sim t_1$, 证毕。

 **练习 1.8** 假设 $n \times 1$ 随机向量 $\mathbf{x} = (x_1, \dots, x_n)^\top$ 服从球对称分布, 证明 $\sqrt{n}\bar{x}/s_x \sim t_{n-1}$, 其中

$$\bar{x} = \sum x_i/n, \quad s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}.$$

证明 设 $H \in \mathbb{R}^{n \times n}$ 为正交矩阵, 满足 $H^\top \mathbf{1} = \sqrt{n}\mathbf{e}_1$, 令 $\mathbf{y} = H\mathbf{x}$, 由于 $\mathbf{x}$ 是球对称分布, 故 $\mathbf{y}$ 仍然服从球对称分布。

注意到:

$$\bar{x} = \frac{1}{n} \mathbf{1}^\top \mathbf{x} = \frac{1}{n} (\sqrt{n}\mathbf{e}_1)^\top \mathbf{y} = \frac{1}{\sqrt{n}} y_1$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=2}^n y_i^2$$

由于 $\mathbf{y}$ 服从球对称分布, 可以表示为 $\mathbf{y} \stackrel{d}{=} \frac{\mathbf{z}}{\|\mathbf{z}\|} \|\mathbf{y}\|$, 其中 $\mathbf{z} \sim N_n(0, I_n)$, 从而:

$$\frac{\sqrt{n}\bar{x}}{s_x} = \frac{y_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n y_i^2}} = \frac{z_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n z_i^2}} \sim t_{n-1}$$

 **练习 1.9** 假设随机向量 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 独立, 其中 $\mathbf{y}$ 服从球对称分布, 记 Pearson 相关系数

$$R_{xy} = \frac{(\mathbf{x} - \mathbf{1}\bar{x})^\top (\mathbf{y} - \mathbf{1}\bar{y})}{\|\mathbf{x} - \mathbf{1}\bar{x}\| \|\mathbf{y} - \mathbf{1}\bar{y}\|},$$

证明

$$\sqrt{n-2} \cdot \frac{R_{xy}}{\sqrt{1-R_{xy}^2}} \sim t_{n-2}.$$

证明 设 $H \in \mathbb{R}^{n \times n}$ 为正交矩阵, 满足 $H^\top \mathbf{1} = \sqrt{n}\mathbf{e}_1$, 则

$$H\mathbf{y} = (\sqrt{n}\bar{y}, \mathbf{w}^\top)^\top,$$

由于 $\mathbf{w}$ 服从 $n-1$ 维的球对称分布, 因此 $\mathbf{v} = \frac{\mathbf{w}}{\|\mathbf{w}\|} \sim U(S^{n-2})$ 。

注意到 $\|\mathbf{y} - \mathbf{1}\bar{y}\|^2 = \mathbf{y}^\top \mathbf{y} - n\bar{y}^2 = \|\mathbf{w}\|^2$, 记 $\mathbf{u} = \frac{\mathbf{x} - \mathbf{1}\bar{x}}{\|\mathbf{x} - \mathbf{1}\bar{x}\|}$, 利用 Lecture3 推论 3(P12), $\mathbf{u}^\top \mathbf{v} \stackrel{d}{=} v_1$, 则:

$$R_{xy} = \frac{(\mathbf{x} - \mathbf{1}\bar{x})^\top (\mathbf{y} - \mathbf{1}\bar{y})}{\|\mathbf{x} - \mathbf{1}\bar{x}\| \cdot \|\mathbf{y} - \mathbf{1}\bar{y}\|} = \frac{\mathbf{u}^\top \mathbf{w}}{\|\mathbf{w}\|} = \mathbf{u}^\top \mathbf{v} \stackrel{d}{=} v_1 \stackrel{d}{=} \frac{z_1}{\|\mathbf{z}\|}$$

其中 $\mathbf{z} \in N_{n-1}(0, I_{n-1})$ 。

故:

$$\sqrt{n-2} \cdot \frac{R_{xy}}{\sqrt{1-R_{xy}^2}} \stackrel{d}{=} \sqrt{n-2} \frac{z_1}{\sqrt{z_2^2 + \dots + z_{n-1}^2}} \sim t_{n-2}.$$